Optimal Feedback Control Rules Sensitive to Controlled Endogenous Risk-Aversion

Dan Protopopescu[1]
Center for the Study of Organizations and Decisions in Economics
Universitat Autònoma de Barcelona
E-mail: *dan.protopopescu@uab.cat*

June 2008

Abstract

The objective of this paper is to correct and improve the results obtained by Van der Ploeg (1984a, 1984b) and utilized in the literature related to feedback stochastic optimal control sensitive to constant exogenous risk-aversion (Karp 1987; Whittle 1989, 1990; Chow 1993, amongst others). More realistic, the proposed approach deals with endogenous risks that are under the control of the decision-maker. It has strong implications on the policy decisions adopted by the decision-maker during the entire planning horizon.

Keywords: Controlled stochastic environment, rational decision-maker, adaptive control, optimal path, feedback optimal strategy, endogenous risk-aversion, dynamic active learning.

JEL Classifications: C51, C61, C91, D81.

1. Introduction and Survey: Nonlinear Versus Linear Behavior

Behavior optimization is very important in any positive economic analysis. The need to solve optimization problems arises in almost all fields of economic inquiry. Various difficulties are encountered by the decision-maker when modelling real economic phenomena. There are multiple sources of uncertainty that he must deal with by using optimal adequate solutions. Complex real world interactions between economy and environment is the main barrier to applied research within the field of economic modelling. In general, the decision model either makes assumptions on how the decision-maker reponds to his environment or derive this behavior from optimality-based considerations.

Uncertainty is an intimate dimension of economics. This generally arises because of inherent difficulties of perception and information processing. We are generally uncertain about the structure of the model, the numerical values of its parameters of interest and the future values of exogenous or random variables. It is well-known the crucial role that the information plays in the decision-making process of individual agents facing uncertainty. Incorporation and judicious use of further prior information into the statistical procedures will produce better estimators. Greater information reduces the environmental complexity, and hence the decision-maker's uncertainty.

In practice, the decision-maker bases his decisions on some body of knowledge. He does not know which state in the future will in fact hold. When that base of knowledge evolves over time, regulatory decisions evolve too. The knowledge can be viewed as future oriented expertise. In dynamic behavior situations, generally the uncertainty is only gradually solved

[1]Financial support from the Spanish Ministry of Education and Science through grant **PTA-2003-02-00877** and **FEDER** is acknowledged.

through time and the decision-maker may be afforded continuously the opportunity to revise his plan of action.

Under uncertainty, the optimal sequence of decisions depends on not only the expected losses, but also the flexibility in terms of availability of future options associated with each decision. Does not matter the preferred decision-maker's policy, the uncertainty will alter this considerably. We recall here some efficient methods in order to reduce the uncertainty, as the control of the future, the increased power of prediction, or by diffusion (Knight 1971).

The decision-maker can utilize the history of the process and develop an approximate model in order to analyze the system behavior. Unless we profit of the particular structure of the problem, it will generally be very expansive to generate information to approximate a large model with enough accuracy. To include this as part of an iteration procedure would be an order of magnitude more complex. Satisfactory approximations are difficult to obtain. A close approximation is sufficient but not necessary for solving the problem. There is a trade-off between the utility of a better approximation and the increase in computational costs which limit our ability to study the model in a data relevant manner. In general, we have a hierarchy of costs for different levels of approximation of the true model. The agent models the main features of the data generating process in a relatively simplified representation (which becomes gradually more complicated if additional data become available) based on observables and related to prior economic theory. The problem is whether this simplification does or does not involve a loss of information. Economic models are only rough approximations of the true data generating process, generally unknown. Because of the measurement error and other random effects, there is considerable uncertainty in determining whether the observed data actually are generated from the true model. How useful are the models based on approximate solutions to optimal behavior, this is a question whose response is given by the various numerical applications.

The stochastic optimal control theory is well developed in the literature (Fleming and Rishel 1975; Kendrick 1981; Karatzas and Shreve 1988, among others). Mathematically speaking, an optimal control problem is concerned with the determination of the best ways to achieve a set of objectives as indexed by a criterion function when the performance is judged over many periods and when the dynamic behavior of the system is subject to a set of constraints. In the terminology of the control theory, the variables are divided into those which represent the condition or state of the objective functional (the so called state variables), and those which guide or control the state variables (the so called control variables).

Depending on the difficulty of the problem, the dynamic optima can be either theoretically analyzed or empirically tested. Improvements in finding closed-form solutions of dynamic stochastic models is still very slow. In order to obtain analytically tractable results, restrictions which are less attractive from an economic point of view have to be imposed.

The only model which can be solved in any generality is the linear-quadratic approximation model which gives linear decision rules under given specific conditions, very convenient for theoretical analyses and attractive on computational grounds.

Linear models are widely used in the literature due to their theoretical simplicity and flexibility or for avoiding hard numerical estimation. One can think of a dynamic model as being linear if its global properties can be completely characterized by its local behavior. Non-linear dynamic models do not have this property of equivalence between local and global dynamics, and thus are substantially more complicated to analyze. The non-linearity typically impedes analytical solutions for an optimization problem. The non-linear modelling is generally less amenable, especially by the presence of uncertainty. This is another reason for which the literature is generally focused on the linear model.

Non-linearity may arise in diverse ways in the econometric applications and there are many possible approaches for specifying non-linear models. One can have, for example, non-linearity in parameters and /or in variables as well as non-linearities in time series or with respect to the system disturbances (AMEMIYA 1974).

In non-linear dynamic models (described by non-smooth functional forms), very special assumptions have to be made in order to obtain closed-form solutions. If the dimension of non-linear models is high, then the analytical treatment becomes very difficult. Moreover, they present considerable difficulties in terms of initialization and convergence. The existence of the optimum as well as the speed of convergence of the algorithm is restricted to certain configurations of the initial parameters of interest. It implies an adjustment mechanism of the tatonnement type.

The deviation caused by the non-linearities in the model is quite important in the sense that while deterministic optimal trajectory follows the desired path quite closely, the stochastic optimal trajectory does not. The more the model is complex, the more difficult will be to track the targets. As flexible as the non-linear dynamic model may be, there is a substantial specification uncertainty. When the analytical formula cannot be obtained easily, the analysis of the problem requires the use of some numerical computational algorithms or simulation techniques-based methods. They remain the only viable way to obtain insights about the system studied. However, difficulties in terms of numerical computation /implementation arise, because the dynamic optimization problem may be characterized by multiple optima.

Confrontation with data is very important. A number of numerical methods have been proposed in the literature of stochastic simulation (TAYLOR AND UHLING 1990; MARCET 1994; AMMAN 1995; RUST 1996; JUDD 1998, AMONG OTHERS). The continue increase in the computers computational speed makes feasible new adaptive control learning algorithms (designed for experimentation) and enlarges the class of models that can be approached by simulation using the data generating process. They are playing an increasingly important role in economic analysis (especially in controlling economic dynamics) and allow to gain experience from large structural models whose properties are revealed by empirical experimentation.

The dynamic programming method provides, in this sense, a constructive recursive procedure for computing the optimal decision rules (BELLMAN 1957; ARIS 1964; BECKMANN 1968; DREYFUS AND LAW 1977; BERTSEKAS AND SHREVE 1978; WHITTLE 1982; ROSS 1983; CAPUZZO, FLEMING AND ZOLEZZI 1985; SNIEDOVICH AND DEKKER 1992, AMONG OTHERS).

This procedure (based on a process of backward induction) amounts to a solution algorithm that allows us to obtain numerical solutions to specific problems as well as analytic characterizations of a wide class of problems. Unfortunately, the amount of computation required to obtain the dynamic programming solution rises exponentially with the number of variables in the model, due to the curse of dimensionality (PITCHFORD 1977).

A weakness of these simulation techniques-based methods is that the properties of the model may depend on a particular specification of the true model, and one may get a distorted picture of the properties. In general, when information is gathered in the real world, the data generating process is not independent and free of noise. The simulation results must not be considered as a perfect substitute of the theory but only as an instrument which may confirm the theoretical results. The transition from theoretical models to empirical models is severely constrained by the quantity and the quality of the available data.

The handling of policy instruments requires information on all states of the system, making the policy rule complicated from the point of view of implementation. It is more difficult to investigate the properties of the optimal policy when one allows for complex history dependence. It is therefore essential to maintain a balance between the desire for a more sophisticated economic model and the need for nominal configuration in terms of computer use. This is what makes the research in this area both difficult and interesting.

In the most practical situations (typically in parametric models), a less complicated model is likely to be preferred if we wish to pursue the accuracy of the estimation or to profit of important analytic advantage. There is a gain of information from the theoretical analysis of the linear model viewed as an a priori specification. It may serve as a good illustrative theoretical example by simplifying the analysis considerably. For example, in the context of a game, the linearity of the model allows for a complete characterization of the set of equilibria. This is not the case in a non-linear model. The assumption of linearity in functional relationships serves to simplify the conceptual and computational development of the theory.

Although linear models are conceptually convenient and analytically flexible, they do not always provide an adequate framework for modelling the economic behavior. More complex models are needed if policy intervention is the purpose of modelling. We need empirical knowledge of the optimal policy performance in different contexts of decision-making. Different contexts call for different actions. The question is whether the differences between linear and non-linear models lead to qualitatively different predictions on the form that learning should take. This problem deserves special attention and many numerical simulations. The linear approach is not preferable when it works with a naive and extremely simplified model.

Efficient dynamic specification tests seem to be relevant for the model selection. In principle, any dynamic optimization model is empirically testable. It allows to study the behavior of the model under different environments. Empirically, there are often conflicts in the criteria of selecting a model to achieve multiple objectives. Several procedures exist for testing the specification of an econometric model in the presence of one or more other models which purport to explain the same phenomenon (Davidson and Mackinnon 1981, among others).

It is useful to note that the rejection of the null hypothesis should not lead to automatic acceptance of the alternative hypothesis, as the test could have a greater power against other deficiencies (Sargan 1988). In other words, the fact that the test fails to reject an hypothesis should not necessarily leads to accept it. The linearity may thus not be rejected if, for example, other variables are added to the initially specified linear model (Granger and Terasvirta 1993). Therefore, it is helpful to know when one can decide that the non-linearity is the element which causes misspecification in the linear model.

It is well-known that misspecified theoretical models could forecast well if the process remains constant, while good models could forecast poorly if the data variance is high. In other words, a model can be acceptable despite having a poor fit and, the fact that a rival model has a better fit does not necessarily make it a better one (Hendry 1995).

Sometimes, it happens that although the model is restrictive and in some ways unrealistic, it brings out many of the key insights. If a model is found to be superior, the matter which remains to be solved is to prove if the difference between the two specifications is significative.

Naturally, all results of the linear model are asymptotically valid in the non-linear specification case. This is the encompassing principle which requires a model to be able to explain characteristics of rival models. However, a model including another does not necessarily encompass this model (Gourieroux and Monfort 1995).

Each econometric approach has its advantages and limitations. New approaches rise new difficulties. The objective of the modeller is to discover the most appropriate model that explains the observed data. In general, there are many implicit restrictions derived from the economic theory. Any specification would be preferred, the builded model must be consistent with the economic theory, data admissible, congruent with the data and computationally attractive (Hendry 1995). There is no royal way to develop good models. There are no precise rules for econometric model design.

The paper is organized as follows. Section 2 deals with the problem statement and makes preliminary considerations. Section 3 presents the model. Section 4 deals with the probabilistic hypotheses on the acquisition of information. Section 5 corrects the results of Van der Ploeg (1984a, 1984b) in the context of a constant exogenous risk-aversion. Section 6 introduces the concept of endogenous risk-aversion. Section 7 improves the formulas obtained in Section 5 by considering endogenous risks that are under the control of the decision-maker. Section 8 draws some conclusions and makes suggestions for further research.

2. Problem Statement and Preliminary Considerations

Facing a risky environment, a rational decision-maker disposes of a set of control instruments in order to constrain the system to follow a fixed optimal trajectory ensuring its equilibrium and stability. The goal is the path. It generally exists a trade-off between the efficiency of control instruments and the decision-maker's objectives in an uncertain and changing world.

The rationality of the decision-maker is characterized by the anticipation that the environment will be affected by other factors than the control instruments. It implies an forward-looking behavior. These factors are completely or partially observed and may be exogenous or endogenous variables. Rationality lies in the correspondence of the decision-maker's action with some goal or objective. He does not refuse to act in accordance with the efficient outcome, at best of his interest (Walsh 1996).

The notion of rational decision-making in an uncertain environment is associated with the expected utility-function maximization behavior. The decision-maker's preferences are generally incomplete. It is very rare in econometrics to be able to fully specify the utility function. No decision-maker has sufficient a priori knowledge to fully specify his preferences.

The decision model is based on the joint use of the econometric model and of the decision-maker's preference function. The latter is optimized under the constraint represented by the former and, very likely, other necessary constraints. The decision will be not separated from the decision procedure and the judgment of rationality carries on the whole.

Learning is one of the three aspects of the decision-maker's uncertainty problem (beside the parametric uncertainty and stochasticity) and has many dimensions. It can take place at various levels of a decision problem. As learning constitutes a form of economic estimation, it is desirable to develop learning algorithms in a context that allows for dynamic structure. Learning possibility can occur only in dynamic models and appears more likely with longer planning horizons. The relative efficiency of the learning generally depends on the method chosen. An optimal behavior may arise from a learning process. However, if the model is very noisy, then the potential for learning is limited. In function of the success of model approximation, the learning may be more or less efficient.

Economic models involving learning often have the potential for converting independent shocks into correlated movements in observables. Models with learning induce persistent effects of transitory shocks. This is an important feature of models with stochastic endogenous fluctuations.

A double learning dynamic is taken into account when analyzing the decision problem of a risk-averse decision-maker: one which describes how the decision-maker adjusts his behavior towards risk over time, and another which reveals the impact of his optimal actions on the system performances. Reinforcement or stimulus-response learning is not generally based on the principle that actions which have led to good outcomes in the past are more likely to be repeated in the future.

The timing of information is a crucial aspect of the decision-making process. The decision-maker can acquire additional information by receiving a noisy signal about the true state of the world. The degree of information embedded in the observation of the state variable generally depends on the values of the control variables, so that the extent of learning about the latent parameters can be influenced directly by the decision-maker. He has some influence over the rate at which information arrives, so that his behavior may generate information. The active learning makes the decision-maker more experienced over time.

In general, the uncertainty will depreciate the decision-maker's activity and will produces a temporary stability followed by a longer or shorter period of adaptation in instability which implies for the decision-maker an additional effort allocated in the active learning.

In a more general context, the instruments can be used for experimenting, the goal being to learn the true parameter of interest. Strategic experimentation is an important aspect of optimal decision-making for a wide class of learning problems. The purpose of experimentation is to gain additional information (which is valuable for future decisions) in order to obtain an optimal learning. The more the decision-maker cares about future performance of the dynamic process, the more he will experiment. The objective of the decision-maker, in this case, is to determine the optimal level of policy experimentation. A rapid decline in the variability of the system state can be associated with an optimal experimentation. It substantially improves the speed of learning as well as the bias in the control and target variables (WIELAND 2000).

Optimal control with learning about unknown parameters has been applied to a variety of economic problems (e.g., optimal investment with production uncertainty, monopolistic pricing with unknown demand, fiscal and monetary policy with imperfect knowledge about the macroeconomy). If the cost of information is too expansive for permitting the learning (e.g., the model is highly non-linear and one searches for the best linear approximation, or the system uncertainties are large), one can be pleased with a rational random behavior (BARBOSA 1975). If the dynamic environment is highly sensitive to non-rational actions, then the stochastic control will be optimal if one can reconcile the desire of the risk with the non-stationarity of the process and the instability of the equilibrium.

In practice, decisions are based on parameters which are not known with certainty and may vary over time. When parameter uncertainty is large, experimentation becomes significatively important. It increases with the variance of the unknown parameters. The degree of experimentation is expected to be smaller with time-varying parameters than with constant ones. In contrast to the constant fixed parameters case (when the incentive to experiment is temporary; it disappears over time as parameter estimates become more precise), the incentive to experiment remains high and never ceases when parameters vary over time (BECK AND WIELAND 2002). Does not matter the type of specification, the incentive to experiment will naturally increase with the variance of random shocks as well as with their degree of persistence.

As regards the decision-maker's strategy, this is based on an adaptive expectation mechanism and on a feedback rule. Control actions adapt as a consequence of changes in endogenous variables and also affect the observability of the system. It will exist feedback between the decision-maker's instruments and the system target variable. This implicitly generates a short-run causality chain. In any discussion of causality, the timing of when things happen is of crucial importance. It must put variables when they occur rather than they are first observed. The decision-maker's actions are taken in real time, whereas his decisions will usually be formulated in advance. In other words, the time passes between taking a decision and its implementation. Sometimes, apparent causality occurs because of the presence of unobserved variables.

The optimality of the strategy adopted is defined relative to the information the decision-maker has at the time the strategy is used. He can use a knowledge base of past and present information to effect a control strategy, but future information is unavailable. Because exogenous shocks in the future are not predictable, the decision-maker strategy cannot incorporate them into the decision. It is assumed that the decision-maker optimally chooses the control instruments on the basis of a non-decreasing endogenous information set.

The control rule is characterized by informational requirements and the decision criterion.

From dynamic economic theory, it is known that optimal decision rules vary systematically with exogenous changes in the structure of series relevant to the decision-maker. It follows that changes in policy will systematically alter the structure of series being forecasted by the decision-maker, and therefore, the behavioral econometric relationships as well. Important cumulative effects of the parameters change on the time path of the state and control variables will be present. It generally exists a relationship between the instruments efficiency and the optimal policy chosen by the decision-maker.

In general, the control process is limited by the speed with which the decision-maker reacts to cautious changes in the environment. There is an inertia of the decision-maker to non-significative environmental changes. This is concretized in deleted observations. The decision-maker generally reacts to sudden shifts of the dynamic system. There is also an inherent inertia effect of the environment due to its capacity of reaction.

At each control period, the level of uncertainty of the decision-maker is given by the deviation of the actual state of the system from his local objective. High deviations from the fixed targets correspond to a high level of uncertainty. The decision-maker adjusts to keep small the difference between actual and assumed system characteristics by monitoring the system fluctuations.

It is assumed that the decision-maker employes a closed-loop strategy, which generally depends on the history of the process, and thus includes feedback information. It may be the case when the relevant information acquisition cost is high, most likely due to the permanent random shocks in the system or because of the slow inertia of the economic environment. This type of strategy has the advantage to continuously improve the decision-maker's optimal policy.

Dynamic feedback entails measurements, and these may be uncertain or indirect. With uncertain or indirect measurements, it is necessary to estimate the state history that is most likely to have caused the measurements. The control principles and the estimation principles are used together to solve the stochastic optimal control problem.

The decision-maker constantly monitors the output of the process under control, the information being employed in real time. The knowledge upon which the decisions are based increases gradually with the passage of time and due to the wisdom derived from experience. The decisions made in the past will be reflected in changes in the state of the system itself and they will influence the perception of the future actions to be analyzed. Because the source of randomness may differ from an application to another, the response of the decision-maker may vary.

In a closed-loop strategy, the policy does not require some large periods of engagement from the part of the decision-maker. In other words, the control rules are sensitive to the choice of the working horizon. Due to the imperfect information about the system reaction over time, it is perfectly reasonable to consider a maximization of short-term for the utility function in the context of a closed-loop strategy.

In general, the length of the working horizon does not only depend on the number of periods but also on the unity of measure chosen. A question remains: What is the optimal length of the planning horizon on which the decision-maker bases his decisions?

An infinite-horizon problem is not generally compatible with a closed-loop strategy. The conceptual and mathematical elegance of infinite horizon models is impractical for a computational viewpoint (even if the policy is easier to implement). To solve such a problem, it is initially convenient to consent ourselves with finite horizon approximates by including some terminal criterion.

The advantage of a finite horizon also lies in the possibility to use forward recursive filtering techniques (Kalman 1960; Kalman and Bucy 1971; Anderson and Moore 1979; Harvey 1990, among others) which allows to monitor the expectation formation process and implicitly the evolution of the stochastic system. This is specific to an incremental learning model based on a sequential forecast which varies with time and history. The infinite horizon problem can be viewed as an approximation of the finite horizon problem with a large planning horizon.

The decision-maker tries to reduce the uncertainty related to the choice of his actions by acquiring information from the beginning of the control to the moment of decision. He has the possibility to learn from errors and to make a self-evaluation of his actions during the period of control.

The closed-loop control is robust in the sense that it anticipates the possibility of a disturbance, and thus can prevent unexpected shocks. This responds not only to the effects of random inputs, but also to the measurement errors as well. It is thus not necessary to be able to identify and measure the sources of disturbance.

If the decision-maker is interested in determining the effect of parameter changes on the optimal control policy, then the closed-loop strategy is generally the best way to do so. The dependence of the coefficients realizations at each point in the horizon is not required in the case of an open-loop approach. For optimal policy experiments and associated hypotheses testing of the optimal control problem, the closed-loop (or feedback) solution is preferable.

The performance of the closed-loop control is superior to any open-loop control in a stochastic dynamic context (CRUZ 1975; XEPAPADEAS 1992; WIEDMER ET AL. 1996, AMONG OTHERS). More the period of control is longer, more the effect of cumulated errors on the decision-maker's optimal policy is significative. If the system evolution is perturbed at each step by a random shock, then the open-loop policy will not integrate this stochastic characteristic for computing the future decisions. The information purchased during the control period is not taken into account, so that the decision-maker will lose the strategic learning. This will affect the optimal policy efficiency as it is adopted on a long-term. The information available to the decision-maker is restricted to the initial value of the state vector. All errors on the initial state of the system will be intactly transmitted until the end of the control process.

The closed-loop strategy is a refinement of the open-loop concept. The open-loop and closed-loop strategies are equivalent only under the perfect forecast assumption, which is unrealistic, in most circumstances. In general, the closed-loop solution deviates from the open-loop solution. Disadvantages of the open-loop controls are that they require much information about the future development of the system and that they are not robust. It is by using a closed-loop strategy that economic theory can be exploited at best.

3. The Model

Consider a stochastic data generating process managed by a system of discrete dynamic simultaneous equations.

Let $x_t \in \mathbf{R}^q$ be the value of the control-related external variable at time t (regarded as a strategic instrument of the agent), let $y_t \in \mathbf{R}^p$ be the system target internal-variable in t (modelled as a partly or indirectly controlled variable), and let $z_t \in \mathbf{R}^r$ be an exogenous variable observed outside the system under consideration, and hence unaffected by the control process at the time period t. It may be forecasted but cannot be influenced by the agent.

Only the inputs and outputs are available to influence and observe the sytem. Hidden state signals are not accessible and may only be estimated using appropriate filters. The control inputs are signals that can be defined arbitrarily by the designer of the control system. The actions are generally dependent variables on the history and current state of the system. For employing the input x_t, the agent will incur a certain cost. Adjustment costs are also incurred for necessary changes in the inputs. In general, the agent is restricted in the use of instruments. Inevitably, there is an arbitrary element in the choice of control variables and an insufficient variability in the instruments.

Whether or not the variable z_t is exogenous depends upon whether or not that variable can be taken as given without losing information for the purpose at hand. Specifically, the exogeneity of the variable z_t depends on the parameters of interest of the agent and on the purpose of the model (statistical inference, forecasting, or policy analysis).

For the purpose of this study, the optimal action x_t is assumed to depend on the observed value of z_t. The variations in the process $\{z_t\}$ will therefore result in variations in the process $\{x_t\}$. Complete learning of the true parameter vector will then depend on whether the process $\{x_t\}$ varies linearly or nonlinearly with the process $\{z_t\}$. Changes in uncertainty about exogenous variables z_t lead to changes in the agent's bias.

In what follows, we make the following basic assumptions:

Assumption 1. The evolution of the system is modelled by the multivariate linear stochastic process:

$$y_t = A_t y_{t-1} + C_t x_t + B_t z_t + D_t + u_t, \qquad t = 1, ..., T$$

where $\beta_t \overset{not.}{=} (A_t,\, C_t,\, B_t,\, D_t) \in \mathbf{R}^k$ is the time-varying parameter to be estimated. It specifies the structure of the model according to the information available in t.

The parameters in the econometric relationships are supposed to vary according to the information accumulated in the system over time. The agent knows that shocks will occur in the future and it is need to be counteracted. Future stabilization of the system is more effective with more precise estimates of the unknown parameters. The whole system is specified and estimated simultaneously.

For optimality reasons, the agent will reestimate the parameters of the model at each period t by taking the feedback effects of learning into account. The process of continuous learning implies an iterative adjustment process and ensures a consistent estimation of the parameters of interest because of the increasingly finer information. The uncertainty on the system parameters is thus renewed at each period. This regular reevaluation of the parameters certifies that the evolution of the estimated model follows that one of the true process. The parameter estimates are only revised in response to forecast errors.

Remark 1. In practice, even if the history of the process is longer, the memory of its states is shorter. The predictable impact of y_{t-1} on y_t depends on the degree of persistence parameter A_t. This is the lagged dependent variable which determines the dynamic of the system. It represents the internal force of the system. The smaller the multiplicative slope parameter B_t, the greater needs to have a compensating moving in the control variable x_t. If the parameter on the control variable is large, then a small change in the control can cause a much larger change in future state. A feature of such control problems is the possibility of a trade-off between current control and estimation.

Note that $u_t \sim i\mathcal{N}(0, \mathbf{\Psi})$ is an exogenous unobserved random shock modelled by a p-dimensional normal distribution with zero mean-vector and finite variance-covariance matrix $\mathbf{\Psi}$. However, nothing forces the data generating process to be stable. The stable distribution is a statistical phenomenon. Note that $\mathbf{\Psi}$ is a non-negative symmetric matrix (not necessarily of full rank) supposed to depend on an unknown nuisance parameter vector which can be either restricted or not. In general, the heteroscedasticity cannot be completely eliminated.

Assumption 2. The agent's objective is to constrain the system to follow a feasible optimal path $\eta \overset{not.}{=} \{y_1^g, y_2^g, ..., y_T^g\}$ by selecting the control variable x_t in a suitable way.

The targets (a priori aspiration levels) reflect the agent's anticipation on the future dynamic of the system, given its backward evolution. Taking into account foreseeable movements in y

as well as possible economic constraints, the agent will fix some optimal bounds l_t such that $0 < y_t^g \leq l_t < 1$, $t = 1, ..., T$.

Assigning extreme values to the targets to be sure that the solution of the model always keeps the values of the objectives on one side of the targets, would not be a realistic strategy for the agent. For stochastic control systems, there are many paths that the system states may follow given the control and initial data. The negative effect of the system stochasticity is the control deviation. The best system performance depends on the information available to the controller at each period t.

Another explanation why the targets are generally unattainable may be their incompatibility with the state of the system. There exist situations when the targets are based on personal assessments rather than on data. The goal of the control is to maintain the process most of its time near the equilibrium state η. An a priori analysis of the deterministic control problem is often crucial (Sargent 1987).

Since a real-time control process is necessarily discrete, this cannot converge with precision to any target value, but only to some neighborhood of it. In other words, after the process of control is ended, the agent will obtain a stochastic neighbouring-trajectory which is expected to be close to the reference-optimal trajectory.

In the real world, it exists permanent and significative errors on the control. Deviations exist because of the phenomenon of "learning bunching", that is, small learning biases are present during some periods while large biases occur during others.

Remark 2. When there is no cost on the control, then the agent does not have as objective to follow a fixed optimal path. This is the case of a myopic (pseudo-optimal) decision behavior.

Remark 3. A necessary condition for the unicity of the instrument is that the number of target variables be inferior to the number of instruments ($p \leq q$).

Assumption 3. The timing of the control is as follows: At each stage t, the agent implements an optimal action x_t, which is a stimulus for the system. This is purported to contribute towards equilibrium and stability. A shock u_t is realized and the agent observes the output y_t (the impulse response) from which he extracts a dynamic signal about the future trend of the system. The agent employes this output signal for a strategic learning (specific to a closed-loop monitoring) in order to drive the system as close as possible to the reference path η. This output and the corresponding action provide information on the data generating process. The uncertainty is reduced only ex-post, that is, only after the informative message has been received. The effect of the shock u_t on the output y_t will disappear gradually in time.

Remark 4. At the end of time period $t - 1$, the control rule x_t is applied and the target variable y_t is determined. The values of x_t and y_t are thus determined at time $t - 1$. During the period $t - 1$ to t the environment reacts to these values, as they become generally known, and by the time t arrives, y_t is determined from the state equation. Consequently, there is an apparent instantaneous relationship between x_t and y_t in the state equation.

The time lag between $t - 1$ and t is a decision lag, and it should be strongly emphasized that this lag need not correspond to the interval between observations of the environment represented by the data available for analysis. Therefore, the decision lag (or decision time) and the observation interval (or observation time-periods) need not coincide, such that one decision period may equal N observation periods (N could be greater than or less than unity).

Assumption 4. The optimality of the instrument x_t is considered with respect to a global criterion which measures the system deviations $\triangle\ y_t \overset{not.}{=} y_t - y_t^g$, $t = 1, ..., T$.

Let $W_{[1,T]}(y_1, y_2, ..., y_T)$ be this criterion, supposed twice continuously differentiable, strictly increasing and convex in the feasible area of the model.

A quadratic objective function may be considered as a good local approximation of the true preferences, exactly as a model approaches the behavior of the system around the observed variables. This is a reasonable one since it induces a high penalty for large deviations of the state variable from the target but a relatively small penalty for small deviations. Note here that no unique criterion unambiguously determines the optimal values for the instruments.

Even in cases where the quadratic criterion is not entirely justified, it is still employed since it leads to an elegant analytical solution for the linear model and a computationally feasible numerical solution for the non-linear model.

Although widely adopted in the literature, the assumption of an explicit expression for the criterion function is not free from critiques, especially when real world problems are considered (Bock and Pauly 1978). There exist situations when the agent is not able or is not willing to formulate an explicit criterion function. How sensitive is this assumption and how sensitive to this assumption are the control results rest still a sensible subject in this area.

In addition, nothing impedes to suppose that the loss function is additively recursive, on one hand, in order to simplify the deduction of the formula for the optimal instruments and, on the other hand, because it makes possible to apply the Bellman (1961) optimality principle backwards through time.

For the purpose of this study, we therefore consider a global criterion which is additive and recursive:

$$W_{[1,T]}(y_1, ..., y_T) \overset{def.}{=} \sum_{t=1}^{T} W_t(y_t)$$

where W_t is a quadratic asymmetric loss function given by:

$$W_t(y_t) \overset{def.}{=} (y_t - y_t^g)' K_t (y_t - y_t^g) + 2(y_t - y_t^g)' d_t$$

with a prime denoting transpose.

The asymmetry of the criterion in the target values derives from the difference in penalty costs that the agent may attach to errors, depending on whether they are errors of shortfall or errors of overshooting about the target. The agent is not thus indifferent with regard to the sign of the system deviations over time. There is an asymmetric treatment of errors to either side of the target. In other words, a positive deviation from a target is not penalized as a negative deviation of the same magnitude.

The criterion for making decisions is a function that puts weight (or measure) on the possible outcomes indicating their desirability or undesirability. The parameters K_t and d_t allow to weight differently the various loss components. In other words, there are not equivalent deviations of the target variables during the optimization process.

The weights used are anything but objective, since the deviation of all target variables may be not of the same importance.

In general, the decision for choosing certain parameters K_t and d_t reflects the agent's priorities and also depends on the available amount of information concerning the future development of the system parameters.

However, it is unlikely that the agent will be able to assign values to the weights which correctly represent his preferences. The idea is to choose the parameters which yield a smoother control (i.e., less fluctuating), and hence a more stable system.

If the future evolution of the system is unpredictable, then the best weighting matrix K_t which can be chosen is the identity matrix, while the best value for the vector d_t is the unity vector.

At each stage t, the parameters K_t and d_t are updated and new optimal values are chosen to satisfy the agent's requirements. These are based on policy values at each stage and do not require any direct information about the actual weighting the agent may have in his mind.

Assumption 5. At each period t, the agent computes his optimal policy $\widehat{x}_t$ before knowing the initial state of the process y_0. He therefore obtains a random optimal policy, conditional to y_0:

$$\widehat{x}_t = \arg\max_{x_t} E_{t-1}[U_t(W_{[1,t]}, \varphi_t) \mid y_0]$$

where $E_{t-1}(\cdot) \overset{not.}{=} E(\cdot \mid I_{t-1})$ is the operator of conditional expectation based on the information available in $t-1$, φ_t is the absolute risk-aversion index at time t, and U_t is the agent's local utility function defined by:

$$U_t(W_{[1,t]}, \varphi_t) \overset{def.}{=} \frac{2}{\varphi_t}[\exp(-\frac{\varphi_t}{2} W_{[1,t]}) - 1]$$

with

$$W_{[1,t]} \overset{def.}{=} \sum_{s=1}^{t} W_s(y_s) \text{ (evolutive loss)}$$

It follows that:

$$-\frac{U_t''(W_{[1,t]}, \varphi_t)}{U_t'(W_{[1,t]}, \varphi_t)} = \frac{\varphi_t}{2}$$

where a prime denotes the partial derivative with respect to $W_{[1,t]}$. Therefore, $\frac{\varphi_t(W_{[1,t]})}{2}$ measures locally (at the point $W_{[1,t]}$) the agent's risk aversion, U_t being a CARA utility. The non-linearity of the utility function is more commonly represented as risk-aversion.

This is Jacobson (1973, 1977) the first who employed an exponential utility for the problems of stochastic optimal control with symmetric quadratic criterion.

Generally speaking, the utility depends on the purposes for which it is developed. It does not exist but for the agent, and thus it has a subjective character. This is derived from individual preferences. The stochastic disturbance in the system will produce random shocks in the agent's preferences over time.

Note that the maximum expected utility solution does not necessarily correspond to a stochastic optimal policy with minimum variance.

Remark 5. It is far from probable that the agent exactly maximizes his utility at each stage of the control. We rather face a nearly optimization behavior, where the control variable is continuously and optimally adjusted to maximize some objective function (Van de Stadt et al. 1985; Varian 1990; Leland 1990, among others).

In general, the initial state y_0 (a past observation of the dynamic process) is either fixed or randomized. In this latter case, the agent can have an a priori distribution on y_0 based on the information acquired up to time $t = 0$.

Because a real system is always subject to permanent shocks, it is not possible to control its initial state exactly. It will amplify the agent's uncertainty on the system behavior. It is crucial to achieve a correct treatment of the starting value y_0 and to measure its impact. Small differences in initial conditions can have large effects on long-run outcomes.

4. Probabilistic Hypotheses on the Acquisition of Information

Given that some random strategies are employed, the stochastic environment must be described by a complete finite probability space $(\Omega, \mathcal{F}, P_\Omega, \mathcal{H})$ endowed with a filtration $\mathcal{H}$ (i.e., an increasing sequence of σ sub-algebras of $\mathcal{F}$) satisfying the usual technical conditions.

Denote by $\mathcal{F}$ the σ-algebra of $\mathcal{P}(\Omega)$. P_Ω is the agent's subjective probability measure on Ω ($\mathcal{P}(\Omega)=1$) and represents the stochastic law of the environment (the agent may be uncertain about the state of the world). In statistical applications, P_Ω is an element of a family of sampling probabilities.

Let $I=\underset{t\leq T}{\cup} I_t$ be the space of all possible "elementary events" in the given environment. It plays the role of Ω. Suppose that the family of events Ω is atomless, that is, that any event but $\emptyset$ (the impossible event) is the union of two exclusive events which are also different from $\emptyset$.

This assumption expresses the idea that a refinement of the description of the uncertain environment can always be made. Additional specific assumptions are also introduced:

Assumption 6. (**Non-anticipation**). The history of the process, the past actions and the history of the exogenous variables constitute the maximum that can be fully observed and known at a given period t.

Assumption 7. (**Non-causality**). The future actions cannot affect the current dynamic of the process.

The principle of causality requires that the dynamics of the process being such that present or past actions can affect only future outcomes and not vice versa.

Assumption 8. (**Retention of information**). At time t, the information I_t is I_{t+1}-measurable. Once the information is obtained, this is definitively acquired. In particular, the past actions are memorized. Uncertainty will be solved over time according to a discrete-filtration $\mathcal{H}\overset{not.}{=}\{\mathcal{F}_t\mid t=0,...,T\}$ with $\mathcal{F}_T\overset{def.}{=}\mathcal{P}(\Omega)$ and $\mathcal{F}_0\overset{def.}{=}\{\emptyset,\Omega\}$ almost trivial (meaning that Ω is the only event of non-zero probability in $\mathcal{F}_0$), filtered to the right with respect to the operator of inclusion (i.e., $\mathcal{F}_t=\underset{s>t}{\cap}\mathcal{F}_s$ for all t, and so $\mathcal{F}_t\subset\mathcal{F}_s$ whenever $s\geq t$). In other words, nothing is forgotten, the memory of the process increasing over time.

5. Linear Feedback Optimal Strategy: The Classical Context

The objective of this section is to correct the theoretical results of Van der Ploeg (1984a, 1984b) for the estimation of the feedback optimal strategy in the context of a linear dynamic stochastic environment.

We consider here the case where the agent's risk-aversion is constant and exogenous by hypothesis. Let φ be the absolute risk-aversion index fixed during the entire control period $[1,T]$.

Proposition 1. Suppose that the matrices $\mathbf{\Psi}^{-1}+\varphi H_t$, $K_t-\varphi H_t(\mathbf{\Psi}^{-1}+\varphi H_t)^{-1}H_t$, and $C_t'[K_t-\varphi H_t(\mathbf{\Psi}^{-1}+\varphi H_t)^{-1}H_t]C_t$ are inversible for each $t=1,...,T$. Under the hypotheses stated in **Section 2** and **Section 3**, the optimal feedback control equation for the period t is given by:

$$\widehat{x}_t(I_{t-1},z_t,\beta_t,K_t,d_t,y_t^g)\mid y_0=G_ty_{t-1}+g_t,\qquad t=1,...,T$$

where:

$$\begin{gathered}G_t=-(C_t'\widetilde{H}_tC_t)^{-1}(C_t'\widetilde{H}_tA_t)\\ g_t=-(C_t'\widetilde{H}_tC_t)^{-1}C_t'[\widetilde{H}_t\,(B_tz_t+D_t)-(I_p-\varphi K_t(\mathbf{\Psi}^{-1}+\varphi H_t)^{-1})h_t]\\ \widetilde{H}_t=K_t-\varphi H_tM_t^{-1}(\varphi)H_t,\quad M_T(\varphi)=\mathbf{\Psi}^{-1}+\varphi H_T\end{gathered}$$

It exists the following backward recurrences ($t=T,T-1,...,1$):

$$\begin{gathered}H_{t-1}=K_{t-1}+(A_t+C_tG_t)'\widetilde{H}_t(A_t+C_tG_t)\\ h_{t-1}=K_{t-1}y_{t-1}^g-(A_t+C_tG_t)'\,[\widetilde{H}_t(C_tg_t+B_tz_t+D_t)-(I_p-\varphi K_t(\mathbf{\Psi}^{-1}+\varphi H_t)^{-1})h_t]\end{gathered}$$

with initial conditions

$$H_T = K_T \text{ and } h_T = K_T y_T^g - d_T$$

Proof. The dynamic programming problem is approached in finite discrete-time and uncertain future. Given the assumptions of non-anticipation, retention of information, and additivity for the global loss function $W_{[1,T]}$, the multiperiod optimization problem (T sub-periods) can be decomposed into a sequence of optimization problems involving only decision variables of each stage, which are easier than the original problem (Bellman 1961):

$$\arg\max_{x_1,...,x_T} E_0 U_T(W_{[1,T]}, \varphi) \;=\; \arg\max_{x_1(\cdot)} E_0[\arg\max_{x_2(\cdot)} E_1[...\arg\max_{x_T(\cdot)} E_{T-1} U_T(W_{[1,T]}, \varphi)]$$

where

$$U_T(W_{[1,T]}, \varphi) \overset{def.}{=} \frac{2}{\varphi}[\exp(-\frac{\varphi}{2}W_{[1,T]}) - 1]$$

represents the agent's utility function at time T and

$$E_{t-1}(\cdot) \overset{not.}{=} E(\cdot \mid I_{t-1}),\; t = 1, ..., T$$

is the operator of conditional expectation based on the information available at time $t-1$.

We deal with a sequential decision problem. It comes to maximize period by period, working every time conditionally to the information acquired. The optimal policy is computed step by step starting from x_T to x_1 (backward through time).

We first consider the decision problem for the last period T, given all the information available at the end of period $T-1$. One can write:

$$E_{T-1}U_T(W_{[1,T]}(y_T), \varphi) = \frac{2}{\varphi}E_{T-1}[\exp(-\frac{\varphi}{2}W_T(y_T))\exp(-\frac{\varphi}{2}\sum_{t=1}^{T-1}W_t(y_t)) - 1]$$

Because the last exponentiel does not depend on x_T, we have:

$$\widehat{x}_T = \arg\max_{x_T} E_{T-1}U_T(W_{[1,T]}(y_T), \varphi) = \arg\max_{x_T} E_{T-1}[\exp(-\frac{\varphi}{2}W_T(y_T))]$$

The assumption of rational expectations makes the problem difficult because the expected value of a non-linear function is not generally the non-linear function of the expected value of the random variable.

Under appropriate regularity conditions, one can interchange the order of integration and differentiation, that is, one can differentiate within the conditional expectation operator.

For the computation of $E_{T-1}[\exp(-\frac{\varphi}{2}W_T(y_T))] \overset{not.}{=} V_T$ (which is supposed to exist), we proceed as follows:

$$E_{T-1}[\exp(-\frac{\varphi}{2}W_T(y_T))] = E_{T-1}[\exp(-\frac{\varphi}{2}(\triangle\, y_T' K_T \triangle\, y_T + 2\triangle\, y_T' d_T))]$$

$$= E_{T-1}[\exp(-\frac{\varphi}{2}(y_T' H_T y_T - 2y_T' h_T + f_T))]$$

where:

$$\triangle\, y_T \overset{not.}{=} y_T - y_T^g,\quad H_T = K_T,\quad h_T = K_T y_T^g - d_T,\quad f_T = y_T^{g\prime}(h_T - d_T)$$

Substituting $A_T y_{T-1} + C_T x_T + B_T z_T + D_T + u_T$ for y_T, one obtains:

$$V_T = E_{T-1}[\exp(-\frac{\varphi}{2}u_T' H_T u_T - \varphi u_T'[K_T(A_T y_{T-1} + C_T x_T + B_T z_T + D_T) - h_T] - \frac{\varphi}{2}f_T)\cdot$$

$$\exp(-\frac{\varphi}{2}((A_T y_{T-1} + C_T x_T + B_T z_T + D_T)'[K_T(A_T y_{T-1} + C_T x_T + B_T z_T + D_T) - 2h_T])))]$$

$$= E_{T-1}[\exp(\omega_2(u_T))] \exp \omega_1(I_{T-1}, x_T, z_T, \beta_T, K_T, d_T, y_T^g)$$

$$= \exp \omega_1(I_{T-1}, x_T, z_T, \beta_T, K_T, d_T, y_T^g) \int_{R^p} (2\pi)^{-\frac{p}{2}} \mid \det \mathbf{\Psi} \mid^{-\frac{1}{2}} \exp(-\frac{1}{2}\widetilde{u}_T' \mathbf{\Psi}^{-1} \widetilde{u}_T) \exp \omega_2(\widetilde{u}_T)\, d\widetilde{u}_T$$

with $\omega_2(\widetilde{u}_T)$ a quadratic function in $\widetilde{u}_T$.
One can write:

$$\widetilde{\mathbf{I}}_T \overset{not.}{=} \int_{R^p} (2\pi)^{-\frac{p}{2}} \mid \det \mathbf{\Psi} \mid^{-\frac{1}{2}} \exp(-\frac{1}{2}\widetilde{u}_T' \mathbf{\Psi}^{-1} \widetilde{u}_T) \exp \omega_2(\widetilde{u}_T)\, d\widetilde{u}_T$$

$$= \int_{R^p} (2\pi)^{-\frac{p}{2}} \mid \det \mathbf{\Psi} \mid^{-\frac{1}{2}} \exp(-\frac{1}{2}\widetilde{u}_T'(\mathbf{\Psi}^{-1} + \varphi H_T)\widetilde{u}_T + \text{linear in } \widetilde{u}_T') d\widetilde{u}_T$$

$$= \mid \det(\mathbf{\Psi}^{-1} + \varphi H_T) \mid^{-\frac{1}{2}} \mid \det \mathbf{\Psi} \mid^{-\frac{1}{2}} \int_{R^p} (2\pi)^{-\frac{p}{2}} \mid \det(\mathbf{\Psi}^{-1} + \varphi H_T) \mid^{\frac{1}{2}} \exp \omega_3(\widetilde{u}_T)\, d\widetilde{u}_T$$

with $\omega_3(\widetilde{u}_T)$ a quadratic function in $\widetilde{u}_T$. Now, we find $\overline{u}_T \in R^p$ such that:

$$\omega_3(\widetilde{u}_T) = -\frac{1}{2}(\widetilde{u}_T - \overline{u}_T)'(\mathbf{\Psi}^{-1} + \varphi H_T)(\widetilde{u}_T - \overline{u}_T) + \text{independent of } \widetilde{u}_T$$

By consequence, we have the following equality:

$$-\frac{1}{2}\widetilde{u}_T'(\mathbf{\Psi}^{-1} + \varphi H_T)\widetilde{u}_T - \varphi \widetilde{u}_T'[K_T(A_T y_{T-1} + C_T x_T + B_T z_T + D_T) - h_T] - \frac{\varphi}{2} f_T$$
$$= -\frac{1}{2}\widetilde{u}_T'(\mathbf{\Psi}^{-1} + \varphi H_T)\widetilde{u}_T + \widetilde{u}_T'(\mathbf{\Psi}^{-1} + \varphi H_T)\overline{u}_T - \frac{1}{2}\overline{u}_T'(\mathbf{\Psi}^{-1} + \varphi H_T)\overline{u}_T$$
$$+ \text{independent of } \widetilde{u}_T$$

It follows that:

$$\text{independent of } \widetilde{u}_T = \frac{1}{2}\overline{u}_T'(\mathbf{\Psi}^{-1} + \varphi H_T)\overline{u}_T - \frac{\varphi}{2} f_T \overset{not.}{=} \omega_4(\overline{u}_T)$$

and

$$-\varphi \widetilde{u}_T'[K_T(A_T y_{T-1} + C_T x_T + B_T z_T + D_T) - h_T] = \widetilde{u}_T'(\mathbf{\Psi}^{-1} + \varphi H_T)\overline{u}_T$$

that is,

$$\overline{u}_T = -\varphi(\mathbf{\Psi}^{-1} + \varphi H_T)^{-1}[K_T(A_T y_{T-1} + C_T x_T + B_T z_T + D_T) - h_T]$$

Therefore, one can write:

$$\widetilde{\mathbf{I}}_T = \mid \det(\mathbf{\Psi}^{-1} + \varphi H_T) \mid^{-\frac{1}{2}} \mid \det \mathbf{\Psi} \mid^{-\frac{1}{2}} \exp(\frac{1}{2}\overline{u}_T'(\mathbf{\Psi}^{-1} + \varphi H_T)\overline{u}_T - \frac{\varphi}{2} f_T) \cdot$$
$$\int_{R^p} (2\pi)^{-\frac{p}{2}} \mid \det(\mathbf{\Psi}^{-1} + \varphi H_T) \mid^{\frac{1}{2}} \exp(-\frac{1}{2}(\widetilde{u}_T - \overline{u}_T)'(\mathbf{\Psi}^{-1} + \varphi H_T)(\widetilde{u}_T - \overline{u}_T)) d\widetilde{u}_T$$

The last integral is equal to 1 because the integrand is the probability density function of a p-dimensional normal random variable:

$$\widetilde{u}_T \sim \mathcal{N}(\overline{u}_T, (\mathbf{\Psi}^{-1} + \varphi H_T)^{-1})$$

with -1 power denoting inverse.

We have the following equality:

$$| \det(\mathbf{\Psi}^{-1} + \varphi H_T) \,|^{-\frac{1}{2}} | \det \mathbf{\Psi} \,|^{-\frac{1}{2}}$$

$$= \quad | \det[\mathbf{\Psi}^{-1}(I_p + \varphi \mathbf{\Psi} H_T)\mathbf{\Psi}] \,|^{-\frac{1}{2}} \quad = \quad | \det (I_p + \varphi \mathbf{\Psi} H_T) \,|^{-\frac{1}{2}}$$

If we replace $\overline{u}_T$ by its value, we find without difficulty:

$$\widetilde{\mathbf{I}}_T = \quad | \det(I_p + \varphi \mathbf{\Psi} H_T) \,|^{-\frac{1}{2}} \exp(-\frac{\varphi}{2}[K_T(A_T y_{T-1} + C_T x_T + B_T z_T + D_T) - h_T]'$$

$$\cdot - \varphi(\mathbf{\Psi}^{-1} + \varphi H_T)^{-1}[K_T(A_T y_{T-1} + C_T x_T + B_T z_T + D_T) - h_T] - \frac{\varphi}{2} f_T)$$

$$= \quad | \det(I_p + \varphi \mathbf{\Psi} H_T) \,|^{-\frac{1}{2}} \exp \omega_4(I_{T-1}, x_T, z_T, \beta_T, K_T, d_T, y_T^g)$$

By consequence, we have:

$$V_T \overset{not.}{=} E_{T-1}[\exp(-\frac{\varphi}{2} W_T(y_T))] = \exp \omega_1(I_{T-1}, x_T, z_T, \beta_T, K_T, d_T, y_T^g) \cdot \widetilde{\mathbf{I}}_T$$

$$= \quad | \det(I_p + \varphi \mathbf{\Psi} H_T) \,|^{-\frac{1}{2}} \exp(\omega_1(I_{T-1}, x_T, z_T, \beta_T, K_T, d_T, y_T^g) + \omega_4(I_{T-1}, x_T, z_T, \beta_T, K_T, d_T, y_T^g))$$

$$= \quad | \det(I_p + \varphi \mathbf{\Psi} H_T) \,|^{-\frac{1}{2}} \exp \omega_5(I_{T-1}, x_T, z_T, \beta_T, K_T, d_T, y_T^g)$$

After several algebraic manipulations, one obtains:

$$\omega_5(I_{T-1}, x_T, z_T, \beta_T, K_T, d_T, y_T^g) = -\frac{\varphi}{2}[y'_{T-1} A'_T \widetilde{H}_T C_T x_T$$

$$+ x'_T C'_T \widetilde{H}_T (A_T y_{T-1} + B_T z_T + D_T) + x'_T C'_T \widetilde{H}_T C_T x_T + (z'_T B'_T + D'_T) \widetilde{H}_T C_T x_T]$$

$$+ \varphi x'_T C'_T [I_p - \varphi K_T (\mathbf{\Psi}^{-1} + \varphi H_T)^{-1}] h_T + \text{independent of } x_T$$

where:

$$\widetilde{H}_T \overset{not.}{=} K_T - \varphi H_T M_T^{-1}(\varphi) H_T$$

$$M_T(\varphi) \overset{not.}{=} \mathbf{\Psi}^{-1} + \varphi H_T = \mathbf{\Psi}^{-1}(\varphi \mathbf{\Psi} + H_T^{-1}) H_T$$

Using the well-known formulas for the derivatives of matrix functions, the first order condition in x_T writes:

$$-\frac{\varphi}{2} C'_T \widetilde{H}_T A_T y_{T-1} - \frac{\varphi}{2} C'_T \widetilde{H}_T (A_T y_{T-1} + B_T z_T + D_T) - \varphi C'_T \widetilde{H}_T C_T x_T$$

$$-\frac{\varphi}{2} C'_T \widetilde{H}_T \,(B_T z_T + D_T) + \varphi C'_T [I_p - \varphi K_T (\mathbf{\Psi}^{-1} + \varphi H_T)^{-1}] h_T = 0 \overset{(\varphi \neq 0)}{\Leftrightarrow}$$

$$-C'_T \widetilde{H}_T A_T y_{T-1} - C'_T \widetilde{H}_T \,(B_T z_T + D_T) + C'_T [I_p - \varphi K_T (\mathbf{\Psi}^{-1} + \varphi H_T)^{-1}] h_T = C'_T \widetilde{H}_T C_T x_T$$

It follows that:

$$\widehat{x}_T(I_{T-1}, z_T, \beta_T, K_T, d_T, y_T^g) = G_T y_{T-1} + g_T \tag{1}$$

$$G_T = -(C'_T \widetilde{H}_T C_T)^{-1} (C'_T \widetilde{H}_T A_T) \tag{2}$$

$$g_T = -(C'_T \widetilde{H}_T C_T)^{-1} C'_T [\widetilde{H}_T \,(B_T z_T + D_T) - (I_p - \varphi K_T (\mathbf{\Psi}^{-1} + \varphi H_T)^{-1}) h_T] \tag{3}$$

The expected utility level for the period T is obtained by substituting for x_T in $\overline{u}_T$:

$$\widehat{V}_T \overset{not.}{=} \quad | \det(I_p + \varphi \mathbf{\Psi} H_T) \,|^{-\frac{1}{2}} \exp \omega_5(I_{T-1}, \widehat{x}_T, z_T, \beta_T, K_T, d_T, y_T^g)$$

$$= \mid \det(I_p + \varphi \mathbf{\Psi} H_T) \mid^{-\frac{1}{2}} \exp(-\frac{\varphi}{2}[y'_{T-1}(A_T + C_T G_T)' \widetilde{H}_T (A_T + C_T G_T) y_{T-1}$$

$$+2y'_{T-1}(A_T + C_T G_T)' \widetilde{H}_T (C_T g_T + B_T z_T + D_T) + (C_T g_T + B_T z_T + D_T)' \widetilde{H}_T (C_T g_T + B_T z_T + D_T)$$
$$-2y'_{T-1}(A_T + C_T G_T)'(I_p - \varphi K_T (\mathbf{\Psi}^{-1} + \varphi H_T)^{-1}) h_T - 2(C_T g_T + B_T z_T + D_T)'$$
$$\cdot (I_p - \varphi K_T (\mathbf{\Psi}^{-1} + \varphi H_T)^{-1}) h_T - \varphi h'_T (\mathbf{\Psi}^{-1} + \varphi H_T)^{-1} h_T + f_T])$$

Now, we include the period $T-1$ in our optimization problem. We have:

$$\widehat{x}_{T-1} \overset{def.}{=} \underset{x_{T-1}}{\arg\max} E_{T-2}[E_{T-1} U_T(W_{[1,T]}(y_T), \varphi)] =$$

$$= \underset{x_{T-1}}{\arg\max} E_{T-2}[E_{T-1}[\exp(-\frac{\varphi}{2} W_T(y_T)) \exp(-\frac{\varphi}{2} \sum_{t=1}^{T-1} W_t(y_t)) - 1]]$$
$$= \underset{x_{T-1}}{\arg\max} E_{T-2}[E_{T-1}[\exp(-\frac{\varphi}{2} W_T(y_T(\widehat{x}_T))) \exp(-\frac{\varphi}{2} W_{T-1}(y_{T-1}))]]$$
$$= \underset{x_{T-1}}{\arg\max} E_{T-2}[\exp(-\frac{\varphi}{2} W_{T-1}(y_{T-1})) E_{T-1}[\exp(-\frac{\varphi}{2} W_T(y_T(\widehat{x}_T)))]]$$

The expected utility level for the two last sub-periods is therefore:

$$V_{T-1} \overset{not.}{=} E_{T-2}[\exp(-\frac{\varphi}{2} W_{T-1}(y_{T-1})) \widehat{V}_T] = \mid \det(I_p + \varphi \mathbf{\Psi} H_T) \mid^{-\frac{1}{2}}$$

$$\cdot E_{T-2}[\exp(-\frac{\varphi}{2}[y'_{T-1} K_{T-1} y_{T-1} - 2y'_{T-1} K_{T-1} y^g_{T-1} + y^{g\prime}_{T-1} K_{T-1} y^g_{T-1}$$
$$+y'_{T-1}(A_T + C_T G_T)' \widetilde{H}_T (A_T + C_T G_T) y_{T-1} + 2y'_{T-1}(A_T + C_T G_T)' \widetilde{H}_T (C_T g_T + B_T z_T + D_T)$$
$$-2y'_{T-1}(A_T + C_T G_T)'(I_p - \varphi K_T (\mathbf{\Psi}^{-1} + \varphi H_T)^{-1}) h_T$$
$$+(C_T g_T + B_T z_T + D_T)' \widetilde{H}_T (C_T g_T + B_T z_T + D_T) - 2(C_T g_T + B_T z_T + D_T)'$$
$$\cdot (I_p - \varphi K_T (\mathbf{\Psi}^{-1} + \varphi H_T)^{-1}) h_T - \varphi h'_T (\mathbf{\Psi}^{-1} + \varphi H_T)^{-1} h_T + f_T])]$$
$$= \mid \det(I_p + \varphi \mathbf{\Psi} H_T) \mid^{-\frac{1}{2}} E_{T-2}[\exp(-\frac{\varphi}{2}[y'_{T-1} H_{T-1} y_{T-1} - 2y'_{T-1} h_{T-1} + f_{T-1}])]$$

where, by identification, one obtains the following recurrence relations:

$$H_{T-1} = K_{T-1} + (A_T + C_T G_T)' \widetilde{H}_T (A_T + C_T G_T)$$
$$h_{T-1} = K_{T-1} y^g_{T-1} - (A_T + C_T G_T)' [\widetilde{H}_T (C_T g_T + B_T z_T + D_T)$$
$$-(I_p - \varphi K_T (\mathbf{\Psi}^{-1} + \varphi H_T)^{-1}) h_T]$$
$$f_{T-1} = y^{g\prime}_{T-1} K_{T-1} y^g_{T-1} + (C_T g_T + B_T z_T + D_T)' \widetilde{H}_T (C_T g_T + B_T z_T + D_T)$$
$$-2(C_T g_T + B_T z_T + D_T)'(I_p - \varphi K_T (\mathbf{\Psi}^{-1} + \varphi H_T)^{-1}) h_T + f_T - \varphi h'_T (\mathbf{\Psi}^{-1} + \varphi H_T)^{-1} h_T$$

The solution for $\widehat{x}_{T-1}$ will be identical with (1) with T replaced by $T-1$, where G_{T-1} and g_{T-1} are defined by (2) and (3), respectively, with a similar change in time subscripts.

One can thus apply a backward induction in time in order to find the agent's optimal strategy for all sub-periods. At the end of this process, one obtains $\widehat{x}_1 = G_1 y_0 + g_1$ as the optimal policy for the first period and the associated maximum expected utility for all periods.

The determination of the optimal $\widehat{x}_1$ depends on the method of forward-looking which is used in the optimality of the future decisions. One cannot obtain an optimal policy for the first period if its behavior in the future is not known.

The matrices G_t are obtained by solving the matrix equations:

$$G_t = -(C_t'\widetilde{H}_t C_t)^{-1}(C_t'\widetilde{H}_t A_t)$$

$$H_{t-1} = K_{t-1} + (A_t + C_t G_t)'\widetilde{H}_t(A_t + C_t G_t)$$

backward in time with initial condition $H_T = K_T$.

The vectors g_t are obtained by solving:

$$g_t = -(C_t'\widetilde{H}_t C_t)^{-1}C_t'[\widetilde{H}_t\,(B_t z_t + D_t) - (I_p - \varphi K_t(\mathbf{\Psi}^{-1} + \varphi H_t)^{-1})h_t]$$

$$h_{t-1} = K_{t-1}y_{t-1}^g - (A_t + C_t G_t)'\,[\widetilde{H}_t(C_t g_t + B_t z_t + D_t) - (I_p - \varphi K_t(\mathbf{\Psi}^{-1} + \varphi H_t)^{-1})h_t]$$

backward in time with initial condition $h_T = K_T y_T^g - d_T$.

These formulas correct the results obtained by Van der Ploeg (1984a, 1984b). In particular, for $\varphi = 0$, one obtains a correction of the results obtained by Chow (1973, 1976a, 1976b, 1977, 1978, 1981, 1993) in the risk-neutral context.

It is possible that differences between ex-ante decisions and ex-post results (i.e., between ex-ante and ex-post optimality) exist. What was in the agent's ex-ante best interest is not necessarily in his ex-post best one.

Even if the linear approximation is only roughly, one can however implement a feedback strategy for a closed-loop dynamic process which is sufficiently good, on one hand for obtaining the evaluation of the policy for the first period and, on the other hand, for the actual implementation in the future.

A comparison between the risk-neutrality and the risk-aversion cases will allow us to analyze the shape of the optimal solution. the risk-neutrality case can then be used as a benchmark, to be compared with several possible values for risk-aversion parameters. Letting the risk-aversion coefficient vary for a given specification of the agent's utility function is not the only way to assess the impact of attitudes towards risk. Another important option is to consider alternative utility functions, characterized by different degrees of absolute risk-aversion. This important aspect will be investigated in the Section 7.

6. Endogenous Risk-Aversion

In the real world, the agent is confronted with multiple risks. His decision is not made independently but jointly with other decisions, which place the agent in risky situations. Decisions made to avoid, even partially, a source of risk may be affected by the presence of others.

It is well-known that economic agents behave on average risk-neutral for small and repeated decisions, but the most common attitude of economic agents in all important decision-making problems is one generated by risk-aversion (they prefer the expected value of the risk to the risk itself). Such a behavior characterizes most decision-makers, at least for large gains or important losses. An agent who expects in the future large deviations from the fixed targets can be considered to be risk-averse.

In general, the aversion is associated with increasing uncertainty while the uncertainty is naturally associated with incomplete information about future behavior of the system. One can interpret the risk as the agent's degree of confidence in the future. It decreases with uncertainty. Traditionally, the risk-aversion is equivalent to the concavity of the agent's utility function or a decreasing marginal utility. However, this is just a way of expressing risk-averse preferences.

In the literature on risk, one generally assume that uncertainty is uniformly distributed over the entire working horizon, when the absolute risk-aversion index is negative and constant. From this perspective, the risk is totally exogenous, and thus independent of endogenous risks.

The classical procedure is "myopic" with regard to potential changes in the future behavior of the agent due to inherent fluctuations of the system over time. The traditional measures of risk-aversion are generally too weak for making comparisons between risky situations.

This can be highlighted in concrete problems in finance and insurance, context for which the ARROW-PRATT measures (in the small) give ambiguous results (ROSS 1981).

We extend the ARROW-PRATT approach (1964, 1971A, 1971B), which takes into account only attitudes towards small exogenous risks, by integrating in the analysis potentially high endogenous risks which are under the control of the agent. This has strong implications on the agent's adaptive behavior in a highly fluctuating environment.

In any uncertain environment, the agent must form expectations. In the case where there is a discrepancy between what the agent expects and reality, his uncertainty will be high. In an noisy environment, the expectations may be disappointed.

The agent can influence the likelihood of the system states by using a reinforcement learning strategy. We say, in this case, that he is not myopic in the sense of expecting. A myopic behavior leads to an important bias in the controls and targets variables.

Future anticipations play an important role in how the agent will decide what strategic actions and optimal risk to take. His behavior depends on forecasts of future system state. Forecasts are updated each time as new observation becomes available.

The agent's rationality is characterized by the fact that the sequence of updated forecasts will converge to the equilibrium of the system. If the data generating process changes in ways not anticipated by the model, then the forecasts lose accuracy. Without uncertainty, the distinction between present and future is confused and there is no anticipation.

Suppose that the agent is a strategic decision-maker. He thinks about the future. Depending on the way the agent perceives future outcomes, both risk sensitivity and optimal decisions will be affected during the process of optimization and control.

Different forecasts are obtained from different information structure. There are several sources of forecast uncertainty, including parameter non-constancy, estimation uncertainty, variable uncertainty, innovation uncertainty and model misspecification.

A correct evaluation of the past is crucial for making optimal predictions in the future. This is necessary for an optimal assessment of the agent's risk aversion. Fluctuations in the system target variable generate a time-varying risk-aversion during the control period.

We make the following useful notations:

$$S_{t,\ p\_d} \overset{not.}{=} \parallel y_{t-1} - y^g_{t-1} \parallel^2 + ... + \parallel y_{t-k_1} - y^g_{t-k_1} \parallel^2$$
(the sum of squared past deviations at time t)

$$S_{t,\ a\_f\_d} \overset{not.}{=} \parallel y^a_{t|I_t} - y^g_t \parallel^2 + ... + \parallel y^a_{t+k_2|I_{t+k_2}} - y^g_{t+k_2} \parallel^2$$
(the sum of squared anticipated future deviations at time t)

$$S_{t,\ w\_p\_d} \overset{not.}{=} \parallel y_{t-1} - y^g_{t-1} \parallel^2 L_{t-1} + ... + \parallel y_{t-k_1} - y^g_{t-k_1} \parallel^2 L_{t-k_1}$$
(the weighted sum of squared past deviations at time t)

$$S_{t,\ w\_a\_f\_d} \overset{not.}{=} \parallel y^a_{t|I_t} - y^g_t \parallel^2 \overline{L}_t + ... + \parallel y^a_{t+k_2|I_{t+k_2}} - y^g_{t+k_2} \parallel^2 \overline{L}_{t+k_2}$$
(the weighted sum of squared anticipated future deviations at time t)

where y^g_{t+i} $(i = 0, ..., k_2)$ represent fixed targets in the future (taking into account foreseeable movements in y), $y^a_{t+i|I_{t+i}}$ $(i = 0, ..., k_2)$ are expected values of the target variable at time $t+i$ based on non-decreasing endogenous information sets I_{t+i} and L_{t-j_1} $(j_1 = 1, ..., k_1)$, $\overline{L}_{t+j_2}$

$(j_2 = 0, ..., k_2)$ are strategic weights attached to the system deviations (in the past and future) with respect to the equilibrium path η.

Econometric forecasting is an useful instrument for the agent. In real decision-making problems, the forecast must be as accurate and efficient as possible. A necessary preliminary step for the agent to optimally choose the target path is to make some a priori expectations on the future evolution of the system based on its past performances.

A question arises: Are the current and past values of the process y_t sufficient to forecast y_{t+j} $(j = 1, ..., k_2)$?

Ex-ante expectations refer to those which held prior to the acquisition of information and generally imply a discrete-time process of tatonnement. They must be unique and in accord with the agent's observations and generally are dependent on the initial value of the state variable.

The more they are distant in time, the more they are difficult to assess (due to the extreme uncertainty of the far future). Ex-ante and ex-post forecast errors are viewed as indicators of uncertainty of the decision-making process.

We are now in a position to give a definition of the agent's risk aversion index by taking into account past performances of the system (a truncated history) and rational anticipations of the system behavior in the future.

Definition. Using t to denote time, the absolute risk-aversion index φ_t evolves according to the following relationship:

$$\varphi_t \overset{def.}{=} \frac{S_{t,\, w\_p\_d} + S_{t,\, w\_a\_f\_d}}{\sqrt{(S_{t,\, p\_d} + S_{t,\, a\_f\_d})^2 + l}},\ t = 1, ..., T$$

where $l \geq 1$ is a fixed integer characterizing the agent's type, and the strategic parameters $L_{t-j_1}, \overline{L}_{t+j_2}$ $(j_1 = 1, ..., k_1;\ j_2 = 0, ..., k_2)$ verify the following inequalities:

$$-1 < L_{t-1} \leq ... \leq L_{t-k_1} \leq 0;\ \ -1 < \overline{L}_t \leq ... \leq \overline{L}_{t+k_2} \leq 0$$

with

$$1 \leq k_1 < T;\ k_2 \geq 0;\ 1 \leq k_1 + k_2 \leq T - 1$$

The weights may differ across individuals. They are updated each time as new observation becomes available. The agent gives a higher importance to the past and future deviations which are closer to the moment of implementation of a new optimal action. Smaller the weight is, higher is the importance given by the agent to the system deviation from his local objective.

Given the potential destabilizing role of a long memory of the process, the agent includes in the analysis only a limited history. Distant past observations might increase significatively the bias of the estimates in the econometric model. In general, these provide an imprecise signal for the agent.

In general, it exists an arbitrary element as regards the choice of the backward lag k_1. The objective is to find the better compromise between fit and complexity. The larger the forward lag k_2 is, the more the prediction error increases. Distant forecasts are difficult to formulate due to unpredictable external disturbances which generally affect the system performance.

It is only by taking into account both, the past and the expected future, that the agent can optimally evaluate the risk in an evolving environment. It allows for a better risk allocation at each period of control. Both objectivity and subjectivity characterizes the agent's risk behavior. Its complexity is given by the changing environment design and the agent's typology.

The higher the degree of risk-aversion at time t, the lower the absolute risk-aversion index φ_t. It may be possible to obtain $\varphi_{t_1} = \varphi_{t_2}$ for $t_1 \neq t_2$, that is, a constant risk-aversion for distinct periods of time. When $\varphi_t = \varphi \neq 0$ for $t = 1, ..., T$, the agent is characterized by a constant risk-aversion during the entire planning horizon. In the case where $\varphi_t \cong 0$, the agent can be considered almost risk-neutral at time t.

The experimental evidence shows that individuals overweight extreme events. Let $\varphi_{\min}$ be an optimal risk-aversion threshold fixed by the agent before starting the control and for the entire working horizon $[1, T]$. The objective is not to exceed this limit threshold. Otherwise, the agent becomes excessively risk-averse for the current control period, being characterized by an extreme pessimism. The optimal threshold $\varphi_{\min}$ is such that it offers the best characterization of the agent's type. The higher (smaller) the agent's risk aversion before starting the control, the smaller (higher) the threshold $\varphi_{\min}$. It is important to distinguish between φ_t and $\varphi_{\min}$. In other words, it must distinguish between local risk-aversion (at time t) and global risk-aversion (over the whole period $[1, T]$). For further details, see PROTOPOPESCU (2007).

7. Linear Optimal Feedback Strategy Sensitive to Controlled Endogenous Risk-Aversion

In this section, we improve the formulas obtained for the optimal feedback control rules in the case of a constant exogenous risk-aversion, by considering the more realistic case of time-varying endogenous risks subjected to the control of the decision-maker.

Proposition 2. Suppose that the matrices $\mathbf{\Psi}^{-1} + \varphi_t H_t$, $K_t - \varphi H_t(\mathbf{\Psi}^{-1} + \varphi_t H_t)^{-1} H_t$, and $C_t'[K_t - \varphi H_t(\mathbf{\Psi}^{-1} + \varphi_t H_t)^{-1} H_t] C_t$ are inversible for each $t = 1, ..., T$. Under the hypotheses stated in **Section 2** and **Section 3**, the linear feedback control equations for a rational decision-maker characterized by endogenous risk-aversion are given by:

$$\widehat{x}_t(I_{t-1}, z_t, \beta_t, K_t, d_t, y_t^g) \mid y_0 = \overline{G}_t y_{t-1} + \overline{g}_t, \qquad t = 1, ..., T$$

with the following optimal reaction coefficients:

$$\begin{gathered}\overline{G}_t = -(C_t'\overline{H}_t C_t)^{-1}(C_t'\overline{H}_t A_t) \\ \overline{g}_t = -(C_t'\overline{H}_t C_t)^{-1} C_t'[\overline{H}_t (B_t z_t + D_t) - (I_p - \varphi_t K_t(\mathbf{\Psi}^{-1} + \varphi_t H_t)^{-1}) h_t] \\ \overline{H}_t = K_t - \varphi_t H_t M_t^{-1}(\varphi_t) H_t, \qquad M_t(\varphi_t) = \mathbf{\Psi}^{-1} + \varphi_t H_t, \qquad h_t = K_t y_t^g - d_t\end{gathered}$$

Proof. If nonconvexities arise in the objective function, it may greatly complicate the search for the optimal control instruments (AMMAN AND KENDRICK 1995) and additional constraints for smoothing and bounding the controls may also be present.

Very often in practice, the optimal policies tend to fluctuate with a large amplitude. The stochastic solutions have a certain dispersion. Instruments variability will generally have a large influence, increasing the error on the targets. The observable fluctuations in instruments are due to the initial impact of the unpredictable shocks and forecast errors.

One remedy to avoid drastic changes from one period to another is to impose preselected upper and lower bounds on the values of the control variables (SANDBLOM AND BANASIK 1985). The bounded control approach with bounds not only on the magnitude but also on the rate of change of the controls holds much benefit in many economic applications. However, this procedure may introduce new sources of error and bias. The bounds on instrument variation will produce truncated distributions, and hence will introduce a bias on instrument variation. Moreover, the variation of instruments is given not by their actual efficiency but by their relative position in the set of available instruments. It will therefore generate discontinuities in the relation between instrument efficiency and optimal policy.

At each period t, the agent will maximize his expected utility function under a set of dynamic constraints imposed in order to avoid drastic changes in the control variable:

$$\arg\max_{x_t} E_{t-1}U_t(W_{[1,t]},\varphi_t)$$

$$s.t. \quad : \begin{cases} y_t = A_t y_{t-1} + C_t x_t + B_t z_t + D_t + u_t \\ \underline{L}'_t \leq x_t \leq \overline{L}'_t & \text{(amplitude bounds)} \\ \underline{L}''_t \leq x_t - x_{t-1} \leq \overline{L}''_t & \text{(change bounds)} \\ y_0,\ y_t > 0, \quad t = 1, ..., T & \text{(economic constraints)} \end{cases}$$

where

$$W_{[1,t]}(y_1,...,y_t) \stackrel{def.}{=} \sum_{s=1}^{t} W_s(y_s)$$

with W_t an asymmetric quadratic loss function, strictly convex and twice continuously differentiable:

$$W_s(y_s) \stackrel{def.}{=} (y_s - y_s^g)'K_s(y_s - y_s^g) + 2(y_s - y_s^g)'d_s = y_s'K_s y_s - 2y_s'h_s + f_s$$
$$h_s = K_s y_s^g - d_s; \qquad f_s = y_s^{g\prime}(h_s - d_s)$$

The first set of constraints is imposed to keep the instruments within specific bounds through time. Possible negative realizations of the instruments are ruled out. The wider the bound on the instrument, the higher the importance given by the agent to the variation of the instrument in that direction, so that they fit to the active learning process.

As regards the last set of constraints, this indicates that the variation of the control variable between two consecutive periods lies within a prespecified bounded interval. The values of this variation can be either positive or negative.

The two sets of constraints are called boundary conditions. They restrict the set of potential optima.

The agent chooses the amplitude /change bounds at each iteration of the control algorithm. He can thus exploit the information on the previous instruments when fixing the bounds for the next instruments, by allowing a greater variability for an efficient instrument rather than for an inefficient one.

The bounds on the instruments are simply the limits up to which the agent decides to extend the research of the optimal solution at each iteration.

Note here that the nonnegativity constraints on the state and control variables are never binding (dependent each other) in an optimal plan (Epstein 1981).

During the period of control, a revision process of the feedback information is required. New information resolves uncertainty over time. The value of the optimal instrument $\widehat{x}_t$ is obtained by a revision of expectations in each previous step of the control. The agent's decisions evolve as result of learning.

Following the reasoning employed in **Proposition 1**, we obtain the analytical formulas for the feedback optimal equations sensitive to controlled endogenous risk-aversion.

One can write:

$$\widehat{x}_t = \arg\max_{x_t} E_{t-1}U_t(W_{[1,t]}(y_t),\varphi_t) = \arg\max_{x_t} E_{t-1}[\exp(-\frac{\varphi_t}{2}W_t(y_t))]$$

where:

$$E_{t-1}[\exp(-\frac{\varphi_t}{2}W_t(y_t))] = E_{t-1}[\exp(-\frac{\varphi_t}{2}(y_t'H_t y_t - 2y_t'h_t + f_t))]$$

with:

$$H_t = K_t, \quad h_t = K_t y_t^g - d_t, \quad f_t = y_t^{g\prime}(h_t - d_t)$$

Substituting $A_t y_{t-1} + C_t x_t + B_t z_t + D_t + u_t$ for y_t, one obtains finally:

$$\overline{V}_t \overset{not.}{=} E_{t-1}[\exp(-\frac{\varphi_t}{2} W_t(y_t))] = E_{t-1}[\exp(\omega_2(u_t))] \exp \omega_1(I_{t-1}, x_t, z_t, \beta_t, K_t, d_t, y_t^g)$$

$$= \exp \omega_1(I_{t-1}, x_t, z_t, \beta_t, K_t, d_t, y_t^g) \int_{R^p} (2\pi)^{-\frac{p}{2}} \mid \det \mathbf{\Psi} \mid^{-\frac{1}{2}} \exp(-\frac{1}{2}\widetilde{u}_t' \mathbf{\Psi}^{-1} \widetilde{u}_t) \exp \omega_2(\widetilde{u}_t)\, d\widetilde{u}_t$$

with $\beta_t \overset{not.}{=} (A_t, B_t, C_t, D_t)$ and $\omega_2(\widetilde{u}_t)$ a quadratic function in $\widetilde{u}_t$.

We have:

$$\widetilde{\mathbf{I}}_t \overset{not.}{=} \int_{R^p} (2\pi)^{-\frac{p}{2}} \mid \det \mathbf{\Psi} \mid^{-\frac{1}{2}} \exp(-\frac{1}{2}\widetilde{u}_t' \mathbf{\Psi}^{-1} \widetilde{u}_t) \exp \omega_2(\widetilde{u}_t)\, d\widetilde{u}_t$$

$$= \mid \det(\mathbf{\Psi}^{-1} + \varphi_t H_t) \mid^{-\frac{1}{2}} \mid \det \mathbf{\Psi} \mid^{-\frac{1}{2}} \int_{R^p} (2\pi)^{-\frac{p}{2}} \mid \det(\mathbf{\Psi}^{-1} + \varphi_t H_t) \mid^{\frac{1}{2}} \exp \omega_3(\widetilde{u}_t)\, d\widetilde{u}_t$$

with $\omega_3(\widetilde{u}_t)$ a quadratic function in $\widetilde{u}_t$. Now, we need $\overline{u}_t \in R^p$ such that:

$$\omega_3(\widetilde{u}_t) = -\frac{1}{2}(\widetilde{u}_t - \overline{u}_t)' (\mathbf{\Psi}^{-1} + \varphi_t H_t)(\widetilde{u}_t - \overline{u}_t) + \text{independent of } \widetilde{u}_t$$

As before, one obtains:

$$\overline{u}_t = -\varphi_t (\mathbf{\Psi}^{-1} + \varphi_t H_t)^{-1} [K_t (A_t y_{t-1} + C_t x_t + B_t z_t + D_t) - h_t]$$

Following the same steps as in **Proposition 1**, the integral becomes:

$$\widetilde{\mathbf{I}}_t = \mid \det(I_p + \varphi_t \mathbf{\Psi} H_t) \mid^{-\frac{1}{2}} \exp \omega_4(I_{t-1}, x_t, z_t, \beta_t, K_t, d_t, y_t^g)$$

We have:

$$\overline{V}_t \overset{not.}{=} E_{t-1}[\exp(-\frac{\varphi_t}{2} W_t(y_t))] = \exp \omega_1(I_{t-1}, x_t, z_t, \beta_t, K_t, d_t, y_t^g) \cdot \widetilde{\mathbf{I}}_t$$

$$= \mid \det(I_p + \varphi_t \mathbf{\Psi} H_t) \mid^{-\frac{1}{2}} \exp \omega_5(I_{t-1}, x_t, z_t, \beta_t, K_t, d_t, y_t^g)$$

with:

$$\omega_5(I_{t-1}, x_t, z_t, \beta_t, K_t, d_t, y_t^g) = -\frac{\varphi_t}{2}[y_{t-1}' A_t' \overline{H}_t C_t x_t + x_t' C_t' \overline{H}_t (A_t y_{t-1} + B_t z_t + D_t)$$

$$+ x_t' C_t' \overline{H}_t C_t x_t + (z_t' B_t' + D_t') \overline{H}_t C_t x_t] + \varphi_t x_t' C_t' [I_p - \varphi_t K_t (\mathbf{\Psi}^{-1} + \varphi_t H_t)^{-1}] h_t + \text{independent of } x_t$$

where:

$$\overline{H}_t \overset{not.}{=} K_t - \varphi_t H_t M_t^{-1}(\varphi_t) H_t$$
$$M_t(\varphi_t) \overset{not.}{=} \mathbf{\Psi}^{-1} + \varphi_t H_t = \mathbf{\Psi}^{-1}(\varphi_t \mathbf{\Psi} + H_t^{-1}) H_t$$

The first order condition in x_t writes:

$$-C_t' \overline{H}_t A_t y_{t-1} - C_t' \overline{H}_t (B_t z_t + D_t) + C_t' [I_p - \varphi_t K_t (\mathbf{\Psi}^{-1} + \varphi_t H_t)^{-1}] h_t = C_t' \overline{H}_t C_t x_t$$

It follows that:

$$\widehat{x}_t(I_{t-1}, z_t, \beta_t, K_t, d_t, y_t^g) \mid y_0 = \overline{G}_t y_{t-1} + \overline{g}_t, \qquad t = 1, ..., T$$

where:

$$\overline{G}_t = -(C_t' \overline{H}_t C_t)^{-1} (C_t' \overline{H}_t A_t)$$

$$\overline{g}_t = -(C_t'\overline{H}_t C_t)^{-1} C_t' [\overline{H}_t (B_t z_t + D_t) - (I_p - \varphi_t K_t (\mathbf{\Psi}^{-1} + \varphi_t H_t)^{-1}) h_t]$$

These formulas improve the results related to stochastic feedback optimal control employed by Karp (1987) and Whittle (1981, 1989, 1990).

The optimal feedback control will stabilize the system because it allows for a free flow of information about the system evolution. The strategic decision rule is reviewed and revised in response to new signals from the environment. The agent will thus refine the distance between the current target position and his fixed objectives. His actions are consistent with the planned objective. The deviations from the targets are minimal amongst all possible deviations because the imperfections on $\widehat{x}_1, ..., \widehat{x}_{t-1}$ will not affect $\widehat{x}_t$. The optimal policy is thus robust to the variance of shocks. The algorithm anticipates future learning when choosing the control for each period, and thus will perturb the system early in time in order to reduce the variance of the parameters estimated later in time. The parameters of the behavioral equation are related to the parameters of both economic environment and objective function. The former are derived from the latter through optimization. Therefore, if the parameters of the economic environment change, the parameters of the behavioral equation will also change. The parameters of the state equation also change if the generating mechanism for x_t changes. Knowledge of the former parameters can be used to derive the parameters of the behavioral equation, which can then be utilized to obtain forecasts of the endogenous target variable.

Note that the sufficient variables for describing $\widehat{x}_t$ belong to some spaces of constant dimension, while the endogenous information set I_t generates a sequence of spaces of increasing dimension.

The existence of the optimum may be restricted to certain configurations of the parameters of interest. Accurate parameters estimates are necessary for an efficient implementation of the agent's optimal policy. They represent a basic information for the learning algorithm, and hence are very important as regards the accuracy of numerical simulations.

Consider first the case where there is an unique optimal solution for each period of control. It is supposed that the agent's objective is to keep the instruments within the interval (0, 3.5).

We give below a graphical illustration in this sense.

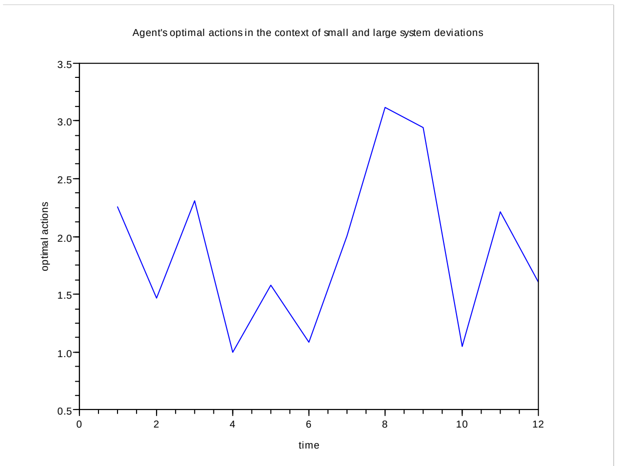

Figure 1

When there is no optimal solution at a given period of time, the agent will choose the most recent value of the control variable which conducted the system close to its optimal target.

Three distinct scenarios may be possible in this particular context:

i) the variable of control takes a negative value, and hence this is ruled out;

ii) the variable of control does not satisfy the condition related to the amplitude bounds;

iii) the variable of control does not satisfy the condition related to the change bounds.

For a numerical illustration, we give below a suggestive graphic for each above scenario.

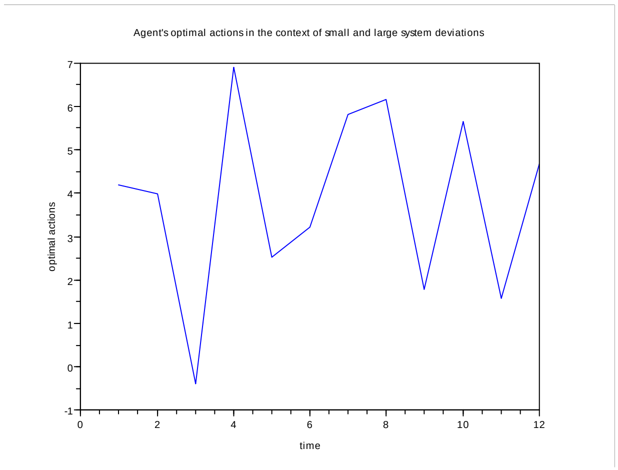

Figure 2-i)

In this case, it is obtained a negative value for $\widehat{x}_3$. The agent will choose between $\widehat{x}_1$ and $\widehat{x}_2$, depending on their performance with respect to the fixed targets y_1^g and y_2^g.

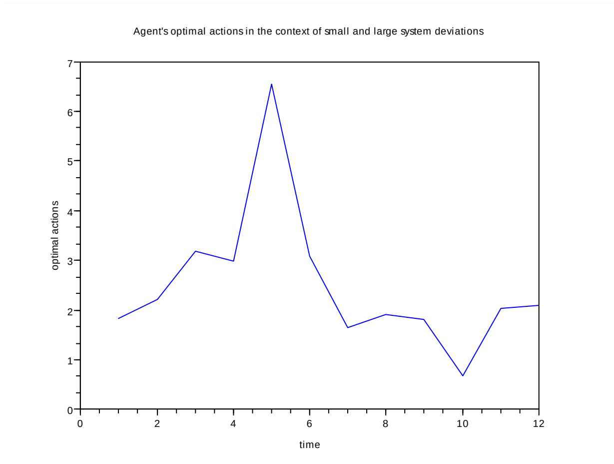

Figure 3-ii)

In this case, the agent's objective is to keep the instruments within the interval $(0, 4)$ during the period of control. At time $t = 5$, the amplitude bound is not satisfied.

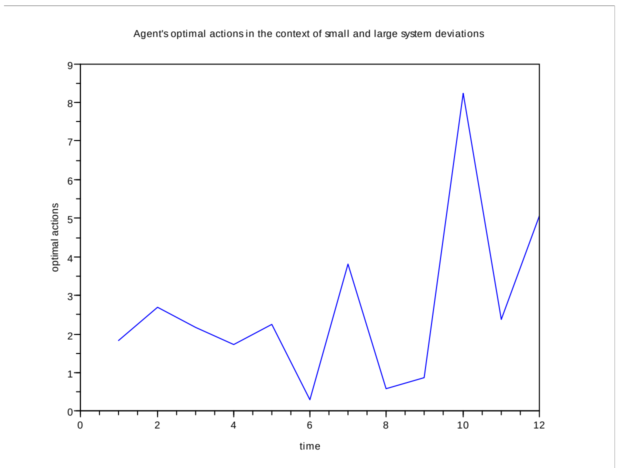

Figure 4-iii)

In this case, the agent's objective is to keep the instruments within the interval $(0, 5]$. Both boundary conditions are not satisfied at time $t = 10$.

It is possible for the agent to implement the same optimal action for distinct periods of time. It may happen if the system is characterized by small fluctuations over time. We illustrate this possibility by a suggestive graphic.

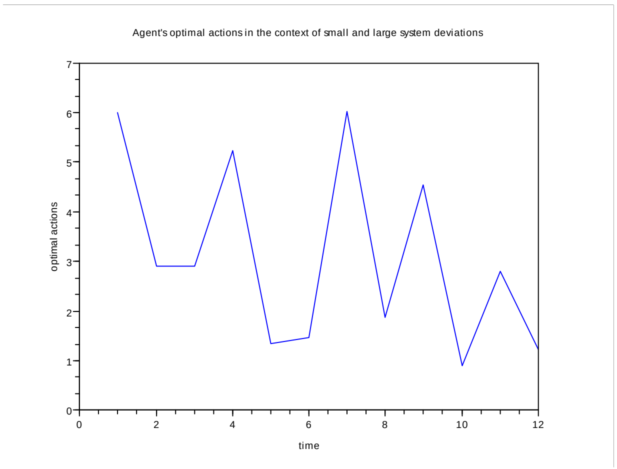

Figure 5

The agent's objective during the entire period of control is to obtain small deviations of the system with respect to the fixed targets. A suggestive graphical illustration is given below.

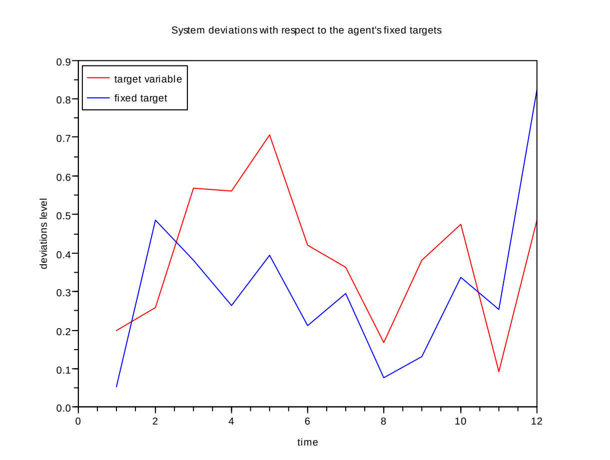

Figure 6

In a noisy environment there should be a trade-off between an increase in information and a decrease in noise. In this particular context, severe problems can be caused to the agent in optimizing the system trajectory. Two distinct scenarios are considered in this sense:

i) the agent does not succeed to constrain the system to follow the optimal trajectory η during the entire planning horizon. We give below a graphical illustration in this sense.

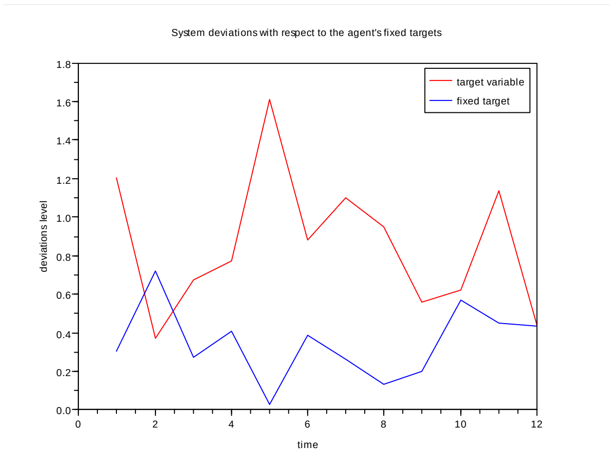

Figure 7

ii) the agent exceeds considerably the fixed targets at each period of the planning horizon.

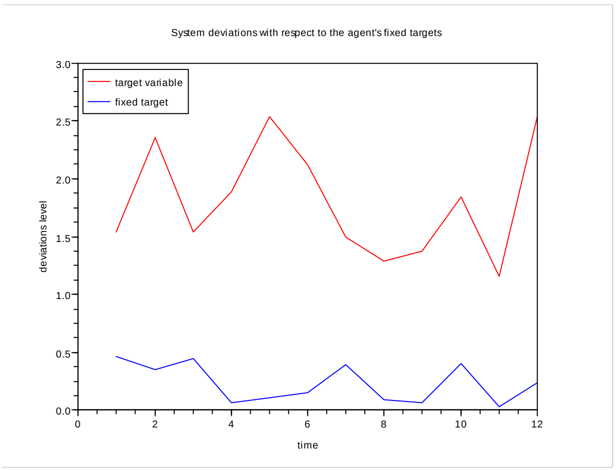

Figure 8

We analyze now the agent's optimal actions with respect to the risk-aversion index level. Two distinct cases are discussed here: i) when the agent is characterized by a small risk-aversion; and ii) when the agent is characterized by a high risk-aversion.

We give below two superposed graphics illustrating these two scenarios.

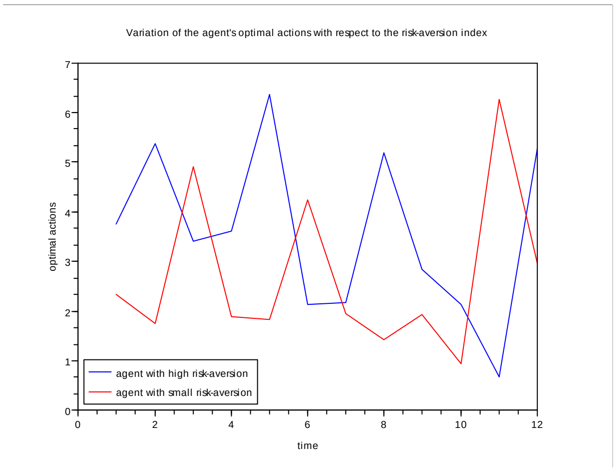

Figure 9

Agent's actions are non-monotonous functions with respect to the risk-aversion index, and hence are not necessarily distinct for distinct attitudes to risk. We examine below the case where the agent's risk aversion index is fluctuating between -1 and 0. Two distinct scenarios are illustrated: i) fluctuating risk-aversion versus risk-neutrality; and ii) fluctuating risk-aversion versus constant risk-aversion.

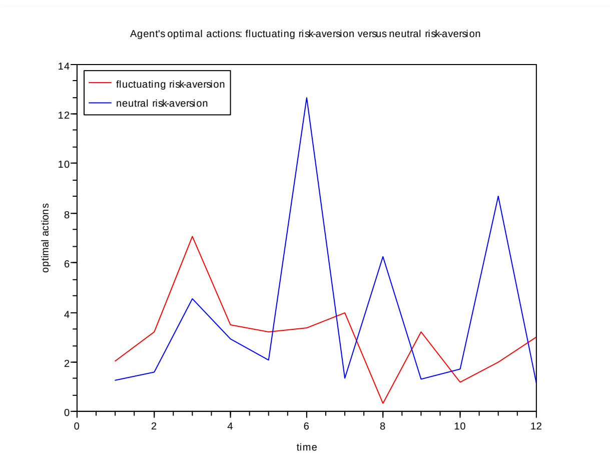

Figure 10-c)

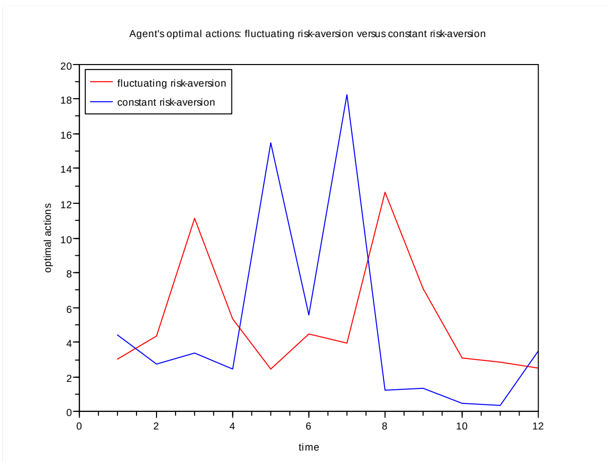

Figure 11-d)

The agent's attitude towards risk is an important element that will condition the shape of the optimal solution. There are important differences between optimal actions sensitive to constant and, respectively, dynamic risk-aversion. It proves the critical importance of the way in which the risk-aversion index is modelled in problems of decision and control, management and planning. Note also that the exceeding of the threshold $\varphi_{\min}$ at time t has a non-negligible effect on the agent's optimal behavior in t. An excessive risk behavior will generate excessive risk actions. It implies either a higher or a lower value of the input x_t. We give below two comparative graphics in this sense.

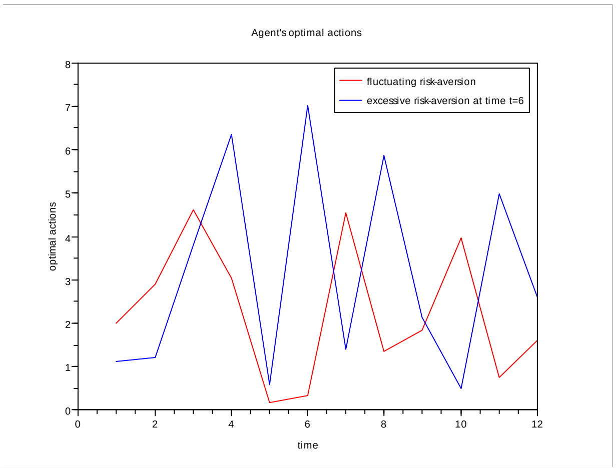

Figure 12

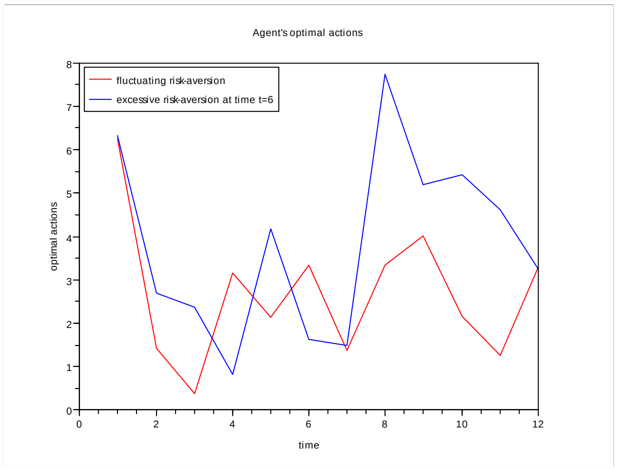

Figure 13

It is important for the agent to optimally choose the weighting parameters K_t and d_t before proceeding at the maximization of his objective function on the regulation horizon. In general, K_t is chosen to be a symmetric positive semi-definite diagonal matrix attaching penalty constant weights to deviations of the state variable from its desired level. If K_t is not diagonal, then penalties also attach to covariances of deviations of the state variable from the desired threshold. The variations of the weighting parameters will affect the extensiveness of the agent's loss function. It is therefore very important to know the effect of K_t and d_t on the agent's risk behavior during the entire control period. The role played by the ponderation matrix K_t can be easily illustrated in the univariate model case. Using the matrix differential rules, one obtains the first-order condition for $\triangle\ y_t$: $2K_t\ \triangle\ y_t + 2d_t = 0 \Leftrightarrow \triangle\ y_t = -\frac{d_t}{K_t}$ (if $K_t \neq 0$). By consequence, if $d_t \neq 0$, each increase (decrease) of the parameter K_t causes an increase (decrease) of the distance between the value of y_t measured and that one fixed at time t. We give below a graphical illustration of the role played by the parameter K_t.

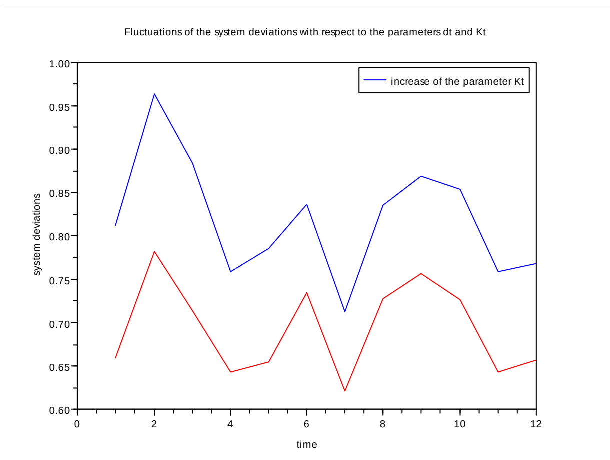

Figure 14

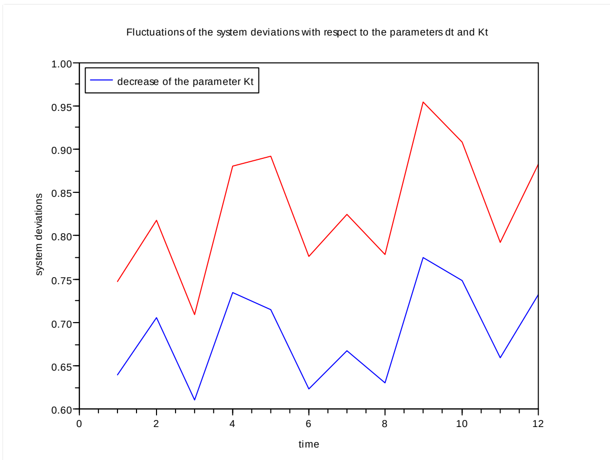

Figure 15

The learning algorithm presented in this paper solves completely the linear-quadratic control problem with suitable initial and boundary conditions in the case of a rational agent characterized by an endogenous risk-aversion during the entire planning horizon. This new approach has the potential to better predict the behavior of the system because integrates a controlled risk-aversion at each period of control. It improves the agent's ability to understand the system response to his implemented actions. In the context of a dynamic game, it will improve the agent's ability to understand a rival's pattern of play. The proposed study also improves the analysis of SAWADA (2008) developed in the context of risk-sensitive tracking control of stochastic systems with preview action.

8. Concluding Remarks and Possible Extensions

This new approach improves previous studies in the literature on control (JACOBSON 1973; KARP 1987; WHITTLE 1981, 1989, 1990; CHOW 1973, 1976A, 1976B, 1977, 1978, 1981, 1993, AMONGST OTHERS) by focusing on optimal feedback control rules sensitive to controlled endogenous risk-aversion. This also corrects the results obtained by VAN DER PLOEG (1984A, 1984B). This work offers to decision-makers (e.g., governments, firms, economic agents) strategic decision rules that allow for a better management and control of dynamic stochastic environments characterized by important fluctuations. Several possible directions for future research can be envisaged here. An immediate application is the case of an asymmetric criterion function expressed as a sum of weighted squares of deviations from given target values for the objectives and instruments. The desired values of the instruments are included in the quadratic loss function to prevent them from going too far away from realistic values. The analysis can also be extended to the case of endogenous targets. The fixed goal is flexible with respect to the possible changes (the nature can change its goal) and can be modified without incurring additional cost, time, or effort. A decision problem is often redefined during the decision process itself. It may be the case where the target path is prescribed without any consideration of the question whether it can be obtained. With endogenous targets, the decision-maker's risk behavior is better shaped. Of great interest is to study the more interesting case of a working horizon that extends as time evolves. The horizon length is thus an endogenous parameter. The resulting moving horizon decision rule will be based on a continuous refinement process of the risk-aversion index. This allows to combine the finite and the infinite horizon optimization problems when the decision-maker exhibits endogenous risk-aversion. Significative differences exist between an individual control problem (viewed as a game against nature), when the decision-maker is submitted only to environment constraints, and respectively a controlled dynamic game, when each player is, in addition, constrained by the opponent's behavior. In this latter case, the equilibrium of the game is subjected to many constraints which mix the parameters of interest. Depending on the nature of the game, we can analyze here two types of behavior: cooperative and non-cooperative. The objective of the players is the optimal risk-sharing during the entire period of the game. We can also test (under heteroskedasticity) if a given discrete time-series arises from a game with Nash /Stackelberg strategies. In the literature, the interest in theoretical and empirical tests aspects in controlled dynamic stochastic games is almost unexistent. We encourage other researchers to take up the challenge.

References

[**1**] AMEMIYA, T., 1974. Non-Linear Regression Models, Chapter **6**, 333-389.

[**2**] AMMAN, H. M., 1995. Numerical methods for linear-quadratic models. In: Amman, H. M., Kendrick, D. A., Rust, J. (Eds.), Handbook of Computational Economics. North-Holland, Amsterdam.

[**3**] AMMAN, H. AND KENDRICK, D., 1995. "Nonconvexities in stochastic control models", *International Economic Review,* **36** (2), 455-475.

[**4**] ANDERSON, B. AND J. MORE, 1979. Optimal filtering (Prentice-Hall, Englewood Cliffs, N J).

[**5**] ARIS, R., 1964. Discrete Dynamic Programming, Blaisdell, New York.

[**6**] ARROW, K. J., 1971A. The Theory of Risk Aversion, Amsterdam, North-Holland.

[**7**] ARROW, K. J., 1971B. Essays in the Theory of Risk Bearing, Chicago, Markham Publishing Co.

[**8**] BARBOSA, F., 1975. "Rational Random behaviour", Doctoral dissertation, *University of Chicago Press.*

[**9**] BECK, G. W AND V. WIELAND, 2002. "Learning and Control in a Changing Economic Environment", *Journal of Economic Dynamics & Control*, **26**, 1359-1377.

[**10**] BECKMANN, M., 1968. Dynamic Programming of Economic Decisions, Springer-Verlag, New York.

[**11**] BELLMAN, R., 1957. Dynamic Programming, *Princeton University Press*, Princeton, New Jersey.

[**12**] BELLMAN, R., 1961. "Adaptive Control Processes: A Guided Tour", *Princetown University Press.*

[**13**] BERTSEKAS, D. P. AND SHREVE, S. E., 1978. "Stochastic Optimal Control: The Discrete Time Case", *Academic Press*, New York.

[**14**] BOCK, V. W. G. AND P. PAULY, 1978. "An optimization approach in multiple target problems using inequality constraints: The case against weighted criterion function", *Annals of Economic and Social Measurement,* **6,** 613-630.

[**15**] CAPUZZO, D. I., FLEMING, W. H. AND ZOLEZZI, T., 1985. Recent Mathematical Models in Dynamic Programming, Springer, Berlin.

[**16**] CHOW, G. C., 1973. "Effect of Uncertainty on Optimal Control Policies", *International Economic Review*, **14**, 632-645.

[**17**] CHOW, G., 1976A. Analysis and Control of Dynamic Economic Systems, John Wiley & Sons, New York.

[**18**] CHOW, G., 1976B. "The Control of Nonlinear Econometric Systems with Unknown Parameters", *Econometrica*, **44**, 685-695.

[**19**] CHOW, G., 1977. "Usefulness of Imperfect Models for the Formulation of Stabilization Policies", *A.E.S.M.*, **6**, 175-187.

[**20**] CHOW, G., 1978. "Evaluation of Macroeconomic Policies by Stochastic Control Techniques", *International Economic Review*, **19**, 311-320.

[**21**] CHOW, G., 1981. Econometric Analysis by Control Methods, John Wiley & Sons, New York.

[**22**] CHOW, G., 1993. "Optimal control without solving the Bellman equation", *Journal of Economic Dynamics & Control* **17**, 621-630.

[**23**] CRUZ, J. B., 1975. "Survey of Nash and Stackelberg Equilibrium Strategies in Dynamics Games", *Annals of Economic and Social Measurement*, **4** (2), 339-344.

[**24**] DAVIDSON, R. AND J. G. MACKINNON, 1981. "Several tests for model specification in the presence of alternative hypotheses", *Econometrica*, **3**, 781-793.

[**25**] DREYFUS, S. E. AND LAW, A. M., 1977. "The Art of Dynamic Programming", *Academic Press*, New York.

[**26**] EPSTEIN, L. G., 1981. "Duality theory and functional forms for dynamic factor demands", *Review of Economic Studies*, **48**, 81-96.

[**27**] FLEMING, W. H. AND R. W. RISHEL, 1975. Deterministic and Stochastic Optimal Control, Springer-Verlag, Berlin.

[28] GOURIEROUX, C. AND A. MONFORT, 1995. "Testing, encompassing and simulating dynamics econometric models", *Econometric Theory,* **11**, 195-228.

[29] GRANGER, J. AND T. TERASVIRTA, 1993. "Modelling Non-linear Economic Relationships", *Oxford University Press.*

[30] HARVEY, A. C., 1990. "Forecasting, Structural Time Series Models and the Kalman Filter", *Cambridge University Press.*

[31] HENDRY, D. F., 1995. "Dynamic Econometrics", *Oxford University Press.*

[32] JACOBSON, D. H., 1973. "Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games", *IEEE Transactions on Automatic Control*, **18**, 124-131.

[33] JACOBSON, D. H., 1977. "Extensions of Linear-Quadratic Control, Optimization and Matrix Theory". London, *Academic Press.*

[34] JUDD, K., 1998. "Numerical Methods in Economics", *MIT Press*, Cambridge, MA.

[35] KALMAN, R. E., 1960. "A New Approach to Linear Filtering and Prediction Problems", *Journal of Basic Engineering,* **82**, **35-45**.

[36] KALMAN, R. E. AND R. S., BUCY, 1971. "New results in Linear Filtering and Prediction Theory", *Journal of Basic Engineering*, **83**, 95-107.

[37] KARATZAS, A. AND S. E. SHREVE, 1988. Brownian motion and stochastic calculus, Springer-Verlag, Berlin.

[38] KARP, L., 1987. "Risk-Aversion in a Dynamic Trading Game", *Working Paper* No. **404**, Division of Agriculture and natural Resources, University of California.

[39] KENDRICK, D. A., 1981. Stochastic control for economic models. Economic handbook Series, McGraw-Hill, New York.

[40] KNIGHT, F. H., 1971. "Risk, uncertainty and profit", *University of Chicago Press*, Chicago, IL.

[41] LELAND, J. W., 1990. A Theory of Approximate Expected Utility Maximization, Mimeo, Department of Social and Decision Sciences, Carnegie Mellon University.

[42] MARCET, A., 1994. "Simulation Analysis of Dynamic Stochastic Models: Application to Theory and Estimation", in Advances in Econometrics, Sixth World Congress. Vol **2**, ed. by C. Sims, Cambridge, *Cambridge University Press.*

[43] PITCHFORD, J. D., 1977. "Two state variable problems", in Pichford, J. D., Turnovsky, S. J., (Eds.), Applications of Control Theory to Economic Analysis, Contributions to Economic Analysis, Vol. **101**, North-Holland, New York.

[44] PRATT, J. W., 1964. "Risk Aversion in the small and in the large", *Econometrica*, **32**, **122-136**.

[45] PROTOPOPESCU, D., 2007. "Improving the Risk Concept: A Revision of Arrow-Pratt Theory in the Context of Controlled Dynamic Stochastic Environments", *Working Paper* No. **727.08**, CODE, Universitat Autònoma de Barcelona, Spain.

[46] ROSS, S. A., 1981. "Some stronger measures of risk aversion in the small and in the large with applications", *Econometrica*, **3**, 621-638.

[47] ROSS, S. M., 1983. "Introduction to Stochastic Dynamic Programming", *Academic Press*, New York.

[48] RUST, J., 1996. "Numerical Dynamic Programming in Economics", in Handbook of Computational Economics, ed. by K. M. Amman, D. A. Kendrick, and J. Rust, Amsterdam, Elsevier Science.

[49] SANDBLOM, C. L. AND J. L. BANASIK, 1985. "Economic policy with bounded controls", *Economic Modelling,* **2**, 135-148.

[50] SANTOS, M. S. AND J. VIGO-AGUIAR, 1998. "Analysis of a numerical dynamic programming algorithm applied to economic models", *Econometrica*, **2**, 409-426.

[51] SARGAN, D., 1988. Lectures on Advanced Econometric Theory, Oxford, Basil Blackwell.

[52] SARGENT, T. J., 1987. "Macroeconomic Theory", *Academic Press*, Orlando.

[53] SAWADA, Y., 2008. "Risk-Sensitive Tracking Control of Stochastic Systems with Preview Action", *International Journal of Innovative Computing, Information and Control*, **1**, 189-198.

[54] SNIEDOVICH, M. AND M. DEKKER, 1992. Dynamic Programming, New York.

[55] TAYLOR, J. B. AND H. UHLING, 1990. "Solving Non-Linear Stochastic Growth Models: A Comparation of Alternative Solution Methods", *Journal of Business and Economic Statistics*, **8**, 1-18.

[56] VAN DER PLOEG, L., 1984A. "Risk and Linear-Quadratic Stabilization", *Economic Letters*, **15**, 73-78.

[57] VAN DER PLOEG, L., 1984B. "Economic Policy Rules for Risk-Sensitive Decision Making", *Zeitschrift für Nationalöekonomie*, **44,** 207-235.

[58] VAN DE STADT, H., KAPTEYN, A., VAN DE GEER, S., 1985. "The Relativity of Utility: Evidence from Panel Data", *Review of Economics and Statistics*, **2**, 179-187.

[59] WALSH, V., 1996. "Rationality, Allocation and Reproduction", *Clarendon Press,* Oxford.

[60] VARIAN, H. L., 1990. "Goodness-of-fit in optimizing models", *Journal of Econometrics,* **46**, 125-140.

[61] WHITTLE, P., 1981. "Risk-sensitive linear-quadratic control problems", *Advances in Applied Probability,* **13**, 764-777.

[62] WHITTLE, P., 1989. "Entropy Minimizing and Risk-Sensitive Control Rules", *System and Control Letters*, **13**, 1-7.

[63] WHITTLE, P., 1990. "Risk Sensitive Optimal Control", Cambridge, *Cambridge University Press.*

[64] WHITTLE, P., 1982. "Optimization Over Time, Dynamic Programming and Stochastic Control", Vol. **1**, Wiley, New York.

[65] WIEDMER, T., FICHER, A. C. AND NAHRAIN, U., 1996. "Pollution Control in a Stochastic Environment", *Swiss Journal of Economics and Statistics*, **132** (4), 575-590.

[66] WIELAND, V., 2000. "Learning by doing and the value of optimal experimentation", *Journal of Economic Dynamics & Control*, **24**, 501-534.

[67] XEPAPADEAS, A. P., 1992. "Environmental Policy Design and Dynamic Non-point Source Pollution", *Journal of Environmental Economics and Management*, **23** (1), 22-39.